NeuroImage 261 (2022) 119509

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

A comparison of methods to harmonize cortical thickness measurements across scanners and sites

Delin Sun[1,2,66], Gopalkumar Rakesh[1,2], Courtney C. Haswell[1,2], Mark Logue[3,4,5,6], C. Lexi Baird[1,2], Erin N. O'Leary[7], Andrew S. Cotton[7], Hong Xie[7], Marijo Tamburrino[7], Tian Chen[7,8], Emily L. Dennis[8,9,10,11], Neda Jahanshad[9], Lauren E. Salminen[9], Sophia I. Thomopoulos[9], Faisal Rashid[9], Christopher R.K. Ching[9], Saskia B.J. Koch[12,13], Jessie L. Frijling[12], Laura Nawijn[12,14], Mirjam van Zuiden[12], Xi Zhu[15,16], Benjamin Suarez-Jimenez[80,15,16], Anika Sierk[17], Henrik Walter[17], Antje Manthey[17], Jennifer S. Stevens[18], Negar Fani[18], Sanne J.H. van Rooij[18], Murray Stein[19], Jessica Bomyea[19], Inga K. Koerte[8,20], Kyle Choi[21], Steven J.A. van der Werff[22,23], Robert R.J.M. Vermeiren[22], Julia Herzog[24], Lauren A.M. Lebois[25,26], Justin T. Baker[27], Elizabeth A. Olson[25,28], Thomas Straube[29], Mayuresh S. Korgaonkar[30], Elpiniki Andrew[31], Ye Zhu[32,33], Gen Li[32,33], Jonathan Ipser[34], Anna R. Hudson[35], Matthew Peverill[36], Kelly Sambrook[37], Evan Gordon[79], Lee Baugh[41,42,43], Gina Forster[41,42,44], Raluca M. Simons[42,45], Jeffrey S. Simons[43,45], Vincent Magnotta[46], Adi Maron-Katz[47], Stefan du Plessis[48], Seth G. Disner[49,50], Nicholas Davenport[49,50], Daniel W. Grupe[51], Jack B. Nitschke[52], Terri A. deRoon-Cassini[53], Jacklynn M. Fitzgerald[54], John H. Krystal[55,56], Ifat Levy[55,56], Miranda Olff[12,57], Dick J. Veltman[58], Li Wang[32,33], Yuval Neria[15,16], Michael D. De Bellis[59], Tanja Jovanovic[60], Judith K. Daniels[61], Martha Shenton[8,62], Nic J.A. van de Wee[22,23], Christian Schmahl[24], Milissa L. Kaufman[25,63], Isabelle M. Rosso[25,28], Scott R. Sponheim[49,50], David Bernd Hofmann[29], Richard A. Bryant[64], Kelene A. Fercho[41,42,43,65], Dan J. Stein[34], Sven C. Mueller[35], Bobak Hosseini[67], K. Luan Phan[67,68], Katie A. McLaughlin[69], Richard J. Davidson[51,52,70], Christine L. Larson[71], Geoffrey May[38,39,40,72], Steven M. Nelson[38,39,40,72], Chadi G. Abdallah[55,56], Hassaan Gomaa[73], Amit Etkin[47,74], Soraya Seedat[48], Ilan Harpaz-Rotem[55,56], Israel Liberzon[75], Theo G.M. van Erp[76,77], Yann Quidé[81,82], Xin Wang[78], Paul M. Thompson[9], Rajendra A. Morey[1,2,*]

[1] *Brain Imaging and Analysis Center, Duke University, Durham, NC, USA.*
[2] *Department of Veteran Affairs (VA) Mid-Atlantic Mental Illness Research, Education and Clinical Center, Durham, NC, USA.*
[3] *National Center for PTSD, VA Boston Healthcare System, Boston, MA, USA.*
[4] *Department of Psychiatry, Boston University School of Medicine, Boston, MA, USA.*
[5] *Biomedical Genetics, Boston University School of Medicine, Boston, MA, USA.*
[6] *Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA.*
[7] *Department of Psychiatry, University of Toledo, Toledo, OH, USA.*
[8] *Psychiatry Neuroimaging Laboratory, Brigham & Women's Hospital, Boston, MA, USA.*
[9] *Imaging Genetics Center, Mark & Mary Stevens Neuroimaging & Informatics Institute, Keck School of Medicine of USC, Marina del Rey, CA, USA.*
[10] *Department of Neurology, University of Utah, Salt Lake City, UT, USA.*
[11] *Stanford Neurodevelopment, Affect, and Psychopathology Laboratory, Stanford, CA, USA.*
[12] *Department of Psychiatry, Amsterdam University Medical Centers, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.*
[13] *Donders Institute for Brain, Cognition and Behavior, Centre for Cognitive Neuroimaging, Radboud University Nijmegen, Nijmegen, The Netherlands.*

[*] Corresponding Author, Rajendra A. Morey, M.D. 40 Duke Medicine Circle, Room 414, Durham, NC 27710 USA, Phone: 919-286-0411 ext. 6425, Facsimile: 919-416-5912,

E-mail address: rajendra.morey@duke.edu (R.A. Morey).

https://doi.org/10.1016/j.neuroimage.2022.119509.
Received 24 August 2021; Received in revised form 7 July 2022; Accepted 22 July 2022
Available online 30 July 2022.

[14] Department of Psychiatry, Amsterdam University Medical Centers, VU University Medical Center, VU University, Amsterdam, The Netherlands.
[15] Department of Psychiatry, Columbia University Medical Center, New York, NY, USA.
[16] New York State Psychiatric Institute, New York, NY, USA.
[17] University Medical Centre Charité, Berlin, Germany.
[18] Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA.
[19] Department of Psychiatry, University of California San Diego, La Jolla, CA, USA.
[20] Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, Ludwig-Maximilians-Universität, Munich, Germany.
[21] Health Services Research Center, University of California, San Diego, La Jolla, CA, USA.
[22] Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands.
[23] Leiden Institute for Brain and Cognition, Leiden, The Netherlands.
[24] Department of Psychosomatic Medicine and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Heidelberg, Germany.
[25] Department of Psychiatry, Harvard Medical School, Boston, MA, USA.
[26] Division of Depression and Anxiety Disorders, McLean Hospital, Belmont, MA, USA.
[27] Institute for Technology in Psychiatry, McLean Hospital, Harvard University, Belmont, MA, USA.
[28] Center for Depression, Anxiety, and Stress Research, McLean Hospital, Belmont, MA, USA.
[29] Institute of Medical Psychology and Systems Neuroscience, University of Münster, Münster, Germany.
[30] Brain Dynamics Centre, Westmead Institute of Medical Research, University of Sydney, Westmead, NSW, Australia.
[31] Department of Psychology, University of Sydney, Westmead, NSW, Australia.
[32] Laboratory for Traumatic Stress Studies, Chinese Academy of Sciences Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China.
[33] Department of Psychology, University of Chinese Academy of Sciences, Beijing, China.
[34] SA MRC Unit on Risk & Resilience in Mental Disorders, Department of Psychiatry and Neuroscience Institute, University of Cape Town, Cape Town, South Africa.
[35] Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium.
[36] Department of Psychology, University of Washington, Seattle, WA, USA.
[37] Department of Radiology, University of Washington, Seattle, WA, USA.
[38] Veterans Integrated Service Network-17 Center of Excellence for Research on Returning War Veterans, Waco, TX, USA.
[39] Department of Psychology and Neuroscience, Baylor University, Waco, TX, USA.
[40] Center for Vital Longevity, School of Behavioral and Brain Sciences, University of Texas at Dallas, Dallas, TX, USA.
[41] Division of Basic Biomedical Sciences, Sanford School of Medicine, University of South Dakota, Vermillion, SD, USA.
[42] Center for Brain and Behavior Research, University of South Dakota, Vermillion, SD, USA.
[43] Sioux Falls VA Health Care System, Sioux Falls, SD, USA.
[44] Brain Health Research Centre, Department of Anatomy, University of Otago, Dunedin, New Zealand.
[45] Department of Psychology, University of South Dakota, Vermillion, SD, USA.
[46] Department of Radiology, Psychiatry, and Biomedical Engineering, University of Iowa, Iowa City, IA, USA.
[47] Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA.
[48] Department of Psychiatry, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa.
[49] Minneapolis VA Health Care System, Minneapolis, MN, USA.
[50] Department of Psychiatry, University of Minnesota, Minneapolis, MN, USA.
[51] Center for Healthy Minds, University of Wisconsin-Madison, Madison, WI, USA.
[52] Department of Psychiatry, University of Wisconsin-Madison, Madison, WI, USA.
[53] Department of Surgery, Division of Trauma and Acute Care Surgery, Medical College of Wisconsin, Milwaukee, WI, USA.
[54] Department of Psychology, Marquette University, Milwaukee, WI, USA.
[55] Division of Clinical Neuroscience, National Center for PTSD, West Haven, CT, USA.
[56] Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA.
[57] ARQ National Psychotrauma Centre, Diemen, The Netherlands.
[58] Department of Psychiatry, Amsterdam University Medical Center, location VUMC, Amsterdam, The Netherlands.
[59] Healthy Childhood Brain Development Developmental Traumatology Research Program, Department of Psychiatry and Behavioral Sciences, Duke University, Durham, NC, USA.
[60] Department of Psychiatry and Behavioral Neurosciences, Wayne State University School of Medicine, Detroit, MI, USA.
[61] Department of Clinical Psychology, University of Groningen, Groningen, The Netherlands.
[62] VA Boston Healthcare System, Brockton Division, Brockton, MA, USA.
[63] Division of Women's Mental Health, McLean Hospital, Belmont, MA, USA.
[64] School of Psychology, University of New South Wales, Sydney, NSW, Australia.
[65] Civil Aerospace Medical Institute, US Federal Aviation Administration, Oklahoma City, OK, USA.
[66] Department of Psychology, The Education University of Hong Kong, Hong Kong, China
[67] Department of Psychiatry, University of Illinois at Chicago, Chicago, IL, USA.
[68] Mental Health Service Line, Jesse Brown VA Chicago Health Care System, Chicago, IL, USA.
[69] Department of Psychology, Harvard University, Cambridge, MA, USA.
[70] Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA.
[71] Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI, USA.
[72] Department of Psychiatry and Behavioral Science, Texas A&M University Health Science Center, Bryan, TX, USA.
[73] Department of Psychiatry and Behavioral Health, Pennsylvania State University, Hershey, PA, USA.
[74] VA Palo Alto Health Care System, Palo Alto, CA, USA.
[75] Department of Psychiatry and Behavioral Science, Texas A&M University, College Station, TX, USA
[76] Clinical Translational Neuroscience Laboratory, Department of Psychiatry and Human Behavior, University of California Irvine, Irvine, CA, USA.
[77] Center for the Neurobiology of Learning and Memory, University of California, Irvine, Irvine, CA, USA.
[78] Department of Mathematics and Statistics, University of Toledo, Toledo, OH, USA.
[79] Department of Radiology, Washington University, St. Louis, MO, USA.
[80] Del Monte Institute for Neuroscience, University of Rochester Medical Center, Rochester, NY, USA.
[81] School of Psychology, The University of New South Wales, Sydney, NSW, Australia.
[82] Neuroscience Research Australia, Randwick, NSW, Australia.

ARTICLE INFO

Keywords:
Data Harmonization
Scanner Effects
Site Effects
Cortical Thickness
ComBat
ComBat-GAM
Linear Mixed-Effects Model
General Additive Model
PTSD

ABSTRACT

Results of neuroimaging datasets aggregated from multiple sites may be biased by site-specific profiles in participants' demographic and clinical characteristics, as well as MRI acquisition protocols and scanning platforms. We compared the impact of four different harmonization methods on results obtained from analyses of cortical thickness data: (1) linear mixed-effects model (LME) that models site-specific random intercepts (LME_{INT}), (2) LME that models both site-specific random intercepts and age-related random slopes ($LME_{INT+SLP}$), (3) ComBat, and (4) ComBat with a generalized additive model (ComBat-GAM). Our test case for comparing harmonization methods was cortical thickness data aggregated from 29 sites, which included 1,340 cases with posttraumatic stress disorder (PTSD) (6.2–81.8 years old) and 2,057 trauma-exposed controls without PTSD (6.3–85.2 years old). We found that, compared to the other data harmonization methods, data processed with ComBat-GAM was more sensitive to the detection of significant case-control differences ($X^2(3) = 63.704$, $p < 0.001$) as well as case-control differences in age-related cortical thinning ($X^2(3) = 12.082$, $p = 0.007$). Both ComBat and ComBat-GAM outperformed LME methods in detecting sex differences ($X^2(3) = 9.114$, $p = 0.028$) in regional cortical thickness. ComBat-GAM also led to stronger estimates of age-related declines in cortical thickness (corrected *p-values* < *0.001*), stronger estimates of case-related cortical thickness reduction (corrected *p-values* < *0.001*), weaker estimates of age-related declines in cortical thickness in cases than controls (corrected *p-values* < *0.001*), stronger estimates of cortical thickness reduction in females than males (corrected *p-values* < *0.001*), and stronger estimates of cortical thickness reduction in females relative to males in cases than controls (corrected *p-values* < *0.001*). Our results support the use of ComBat-GAM to minimize confounds and increase statistical power when harmonizing data with non-linear effects, and the use of either ComBat or ComBat-GAM for harmonizing data with linear effects.

1. Introduction

Large consortia, such as Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) (Thompson et al., 2020), Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) (Hofer et al., 2020), and others have aggregated neuroimaging data acquired on many different scanners and recruited subjects at many different sites to conduct meta- and mega-analyses. By applying standardized analysis pipelines to extremely large datasets of thousands or tens of thousands of samples, consortia improve reliability, enhance reproducibility of results, amass sufficient statistical power to detect relatively small effect sizes, and support the ability to divide samples while retaining the power to delineate subsample (e.g., male vs female or young vs old) and interaction effects. The diverse ethnic, racial, geographic, and clinical demography of consortium data has provided results that are more representative of the wider population while also permitting exploration of clinical and neurobiological subtypes of neuropsychiatric disorders (Dennis et al., 2022; Thompson et al., 2020). Neuroimaging results generated by consortia are more robust and reproducible than studies that are generated by a single laboratory (Koshiyama et al., 2022), provided that consortia apply uniform methods to data originating from multiple sites and scanners.

However, several challenges are posed by the analysis of consortium data. A major concern of consortium-generated results is bias introduced by site-specific acquisition protocols and MRI scanners that may interact with site-specific demographic and clinical profiles (Radua et al., 2020).The challenge of *post hoc* combination of datasets stems partly from a lack of *a priori* harmonization of MRI acquisition sequences. Prospective data collection by consortia such as NCANDA (Brown et al., 2015), ABCD (Volkow et al., 2018), TRACK-TBI (Hicks et al., 2013), and others have prescribed harmonized acquisition parameters at study outset with the expectation of superior performance and obviating the need for post-acquisition harmonization. However, even prospective standardization and prescription of acquisition parameters results in significant variance attributed to sites for relatively short scan duration (e.g., 5 min) that can be reduced significantly by increasing scan duration (e.g., 25 min) (Noble et al., 2017). It remains unclear whether further post hoc harmonization of these datasets may improve sensitivity and power of analyses.

Various methods to harmonize neuroimaging data across sites are gaining acceptance and will become commonplace. However, there is little empirical evidence to support the use of a single method due to the lack of formal comparisons of available methods. In this study, we compared four harmonization methods. First, we tested linear mixed-effects modeling (LME), also known as the mixed-effects mega-analysis (MEMega) (Radua et al., 2020), with site as a random intercept (LME_{INT}) to model the intercept location effects of site on brain measures. Second, we tested LME with both random intercept and age-related random slope for the site covariate ($LME_{INT+SLP}$). Third, we used ComBat, a method originally developed to minimize batch effects present in data originating from multiple gene arrays (Johnson et al., 2007), and later adapted for neuroimaging data. ComBat is designed to remove site-associated differences while preserving variation due to biologically relevant variables such as age, sex, and diagnosis (Fortin et al., 2018). ComBat has been widely used to harmonize neuroimaging data including cortical thickness (Fortin et al., 2018), surface area, subcortical volumes (Radua et al., 2020), diffusion tensor imaging (Fortin et al., 2017; Hatton et al., 2020), and resting-state functional connectivity (Yu et al., 2018). Radua et al. (2020) reported that ComBat and LME_{INT} produced similar results when harmonizing cortical thickness, surface area, and subcortical volumes, while ComBat harmonization led to slightly higher statistical significance when performing between-group comparisons, in a multisite imaging study of schizophrenia. The fourth method, by Pomponio et al. (2020), improves on ComBat by modeling non-linear effects of age with a generalized additive model (GAM). ComBat-GAM allows for varied distributions of scale (multiplicative, or variance) and location (additive, or mean) effects, respectively.

ComBat-GAM was designed to capture age-related non-linearities across the lifespan by fitting a GAM with a penalized nonlinear term. Pomponio et al. (2020) examined cortical and subcortical gray matter volumes without harmonization, harmonized by ComBat, and harmonized by ComBat-GAM in a large sample of 10,477 healthy subjects aggregated from 18 sites who ranged in age from 3 to 96 years. They reported that gray matter volumes harmonized by ComBat-GAM achieved the best performance in an age prediction task that minimized the difference between actual age and predicted age. They also found that ComBat-GAM, compared to other approaches, consistently led to improved prediction accuracy for each dataset in a leave-one-site-out validation experiment. However, Pomponio et al. (2020) only investigated data from healthy participants, which did not involve case-control comparisons, nor formal comparisons to LME methods.

Consequently, the goals of the present study were to investigate (1) the performance of ComBat-GAM for comparing clinical cases to controls, (2) how performance is influenced by age, and (3) how well per-

formance characteristics compare to LME_{INT}, $LME_{INT+SLP}$, and ComBat. Although the random-effects meta-analysis (RE-Meta) has been widely used by ENIGMA projects (Zugman et al., 2022), we did not include RE-Meta in this study because several studies showed that LME and ComBat produce results with greater statistical power than RE-Meta (Boedhoe et al., 2017; Favre et al., 2019; Radua et al., 2020; van Rooij et al., 2018). The increase in power is based on the premise that the site effect being removed represents random noise, and its removal leads to larger effect sizes and greater efficiency requiring fewer subjects to reject the null hypothesis at a pre-specified power.

An important caveat is that performance was measured by the number of brain regions with significant case-control differences. We recognize that neither the method with the greatest number of regions reaching significance nor the method that maximizes the magnitude (absolute value) of regression coefficients reflects the true underlying cortical thickness - the so-called *ground* truth. However, harmonization can move the values further from the ground truth and still be useful. The main aim of harmonization is to make uncalibrated measurements more comparable to each other. It is possible that measurable differences between cases and controls are potentially masked by scanner bias and effective harmonization should increase the difference between the distribution of cases and controls. Therefore, it is advisable to count the number of regions that are statistically significant after implementing harmonization. Nonetheless, there is a risk that harmonization may introduce variability that was not present in the original data.

Data aggregated from 29 sites served as our test case for comparing harmonization methods. Subjects' data was grouped into *cases* with PTSD ($N = 1340$) and trauma-exposed *controls* without PTSD ($N = 2057$). PTSD is associated with anatomical and functional alterations in widely distributed regions of the brain (Dennis et al., 2022; Logue et al., 2018; Wang et al., 2021). Military service members with PTSD and comorbid mild traumatic brain injury (mTBI) experience faster age-associated decline in cortical thickness than controls (Santhanam et al., 2019; Savjani et al., 2017). We hypothesized significant case-control differences in cortical thickness and age-related cortical thinning would be detectable in more brain regions by utilizing ComBat-GAM relative to LME_{INT}, $LME_{INT+SLP}$, and ComBat.

2. Methods

2.1. Participants

Data were obtained for secondary analysis from the ENIGMA-PGC PTSD Working Group. The dataset originated from 29 sites located on five continents (PTSD, $N = 1340$; Trauma-Exposed Controls, $N = 2057$) from a broad age group (6.2–85.2 years old). Three sites were the source of all children and adolescents (Duke De Bellis 9.9 ± 2.5; Leiden University 16.0 ± 1.9; University of Washington 13.2 ± 2.9) and one site was the source of older participants (ADNI-DoD 67.9 ± 3.6), with minimal overlap between the 3 sites with participants under 20 years and sites with participants over 20 years. Only one site contributed both children (Duke University-De Bellis) and adults (Duke University-Morey). Demographic information is summarized in **Table 1**. Clinical measures and assessment of PTSD are explained in the **Supplementary Table S1**. The scanner information is listed in **Supplementary Table S2**. All study sites obtained approval from local institutional review boards or ethics committees. All participants provided written informed consent. Data is available upon request from the corresponding author.

2.2. Imaging data preprocessing

Anatomical brain images were preprocessed at Duke University through a standardized neuroimaging and QC pipeline developed by the ENIGMA Consortium (http://enigma.ini.usc.edu/protocols/imaging-protocols/) (Logue et al., 2018). Cortical thickness measurements were generated using the FreeSurfer software (https://surfer.nmr.mgh.harvard.edu) based on the Destrieux atlas (Destrieux et al., 2010) that contains 74 regions per hemisphere. All sites used FreeSurfer 5.3 for parcellation except ADNI-DoD, Minneapolis VA, and the Waco VA, which used FreeSurfer 6.0, as well as Amsterdam Medical Center and University of South Dakota, which used FreeSurfer 7.1.1 (**Supplementary Table S2**). Briefly, white matter surfaces were deformed toward the gray matter boundary at each surface vertex. Cortical thickness was calculated based on the average distance between the parcellated portions of white and pial surfaces within each region per participant. In each region, any missing value was replaced by the mean cortical thickness averaged across same group of participants (either PTSD or trauma-exposed controls) at the same site

2.3. ComBat harmonization

ComBat removes the effects of site while preserving inherent biological variance in the data (Fortin et al., 2018). In the present study, PTSD diagnosis, age, and sex were designated as biological variables. The ComBat approach was implemented using R scripts (https://github.com/Jfortin1/ComBatHarmonization) running on RStudio (ver. 1.3.1073) and R (ver. 4.0.2). Unlike implementations of LME models that merge data harmonization and statistical analyses, ComBat and ComBat-GAM perform only harmonization and make harmonized data available to the user.

2.4. ComBat-GAM harmonization

PTSD diagnosis, age, and sex were designated as biological variables, and age was specified as the only smooth term in the model. We employed the default setting so that the empirical Bayes estimates were used for site effects, and there were no custom boundaries for the smoothing terms. The ComBat-GAM approach was implemented using Python (ver. 3.8.5) scripts (https://github.com/rpomponio/neuroHarmonize).

2.5. Distribution of non-harmonized, ComBat harmonized, and ComBat-GAM harmonized data

Pairwise comparisons of non-harmonized, ComBat harmonized, and ComBat-GAM harmonized data using the function *pairs()* (from the R package *emmeans*) were applied to the absolute differences between the site-specific mean values and the mean value averaged across sites. The absolute, but not signed values, of the differences were investigated in order to test whether ComBat and ComBat-GAM harmonization led to more consistent distributions Specifically, smaller differences between the site-specific mean values and the mean across sites). The pairwise comparisons were also applied to site-specific standard deviations for cortical thickness across cortical regions. The p-values were adjusted using Bonferroni correction for three pairwise comparisons (i.e., ComBat vs. non-harmonized, ComBat-GAM vs. non-harmonized, ComBat-GAM vs. ComBat). The effects of harmonization by LME models cannot be observed directly because data harmonization and statistical analyses are inseparable in LME methods.

2.6. Statistical models

In all models, we included sex, age, and PTSD diagnosis as fixed factors to estimate their effects on regional cortical thickness, and as covariates for testing interaction effects of interest. Either age by diagnosis interaction, or sex by diagnosis interaction was included in the models as a fixed factor when the corresponding interaction was of interest. The supplementary materials report on the influence of age^2 as a fixed-factor to estimate effects on regional cortical thickness, and for testing interaction effects. Linear modeling was used to analyze data harmonized by ComBat and data harmonized by ComBat-GAM. Cortical thickness data without harmonization was entered into the LME models. The LME_{INT}

Table 1
Demographic information per study site.

	Control			PTSD		
SiteName	*N*	*Female: Male*	*Age (yrs: mean± SD)*	*N*	*Female: Male*	*Age (yrs: mean± SD)*
ADNIDoD[a]	105	1:104	70.0±5.2	80	0:80	67.9±3.6
Amsterdam Medical Center	37	18:19	39.6±10.0	38	17:21	40.4±9.9
Columbia University	35	23:12	35.2±10.6	53	34:19	36.3±9.3
Duke University-De Bellis	86	47:39	10.5±2.6	29	15:14	9.9±2.5
Duke University-Morey	270	59:211	39.7±10.1	114	16:98	40.7±9.9
Ghent University	59	59:0	37.7±12.3	8	8:0	32.6±10.3
University of Groningen	-	-	-	40	40:0	38.2±9.7
U.W. Madison-Grupe	38	1:37	30.7±6.6	19	3:16	30.4±6.2
Emory University-GTP	108	103:5	40.8±12.2	66	66:0	37.0±12.3
INTRuST[b]	254	121:133	34.8±13.0	104	23:81	38.6±10.6
U.W. Milwaukee-Larson	45	23:22	35.5±11.4	19	10:9	29.2±8.6
Leiden University	30	26:4	14.7±1.6	22	19:3	16.0±1.9
University of Mannheim	-	-	-	48	48:0	35.9±11.8
Harvard University-McLean	13	13:0	35.6±10.5	39	39:0	38.2±12.9
Minneapolis V.A.-Disner	95	6:89	33.2±8.6	74	2:72	32.0±7.6
University of Münster	26	21:5	26.5±7.4	21	21:0	27.4±7.0
University of Illinois-Chicago	20	0:20	34.0±8.9	23	0:23	31.3±9.3
Harvard University-Rosso	85	44:41	33.5±9.3	20	13:7	35.3±7.9
University of South Dakota	44	7:37	29.9±6.9	78	17:61	28.8±7.1
Stanford University	1	0:1	61.0±0	68	40:28	36.9±10.3
Stellenbosch University	138	100:38	42.9±14.3	120	87:33	39.4±11.0
University of Toledo	61	27:34	34.3±11.6	15	7:8	40.9±9.5
UCAS-Beijing	36	17:19	48.2±6.8	34	21:13	51.0±6.7
University of Cape Town	55	55:0	28.7±6.4	7	7:0	30.5±7.2
University of Sydney-Westmead	107	71:36	40.4±13.2	48	25:23	39.0±11.6
University of Washington	202	105:97	14.1±3.0	53	25:28	13.2±2.9
Waco V.A.	25	4:21	40.7±11.6	41	6:35	41.0±11.0
University of New Haven	34	3:31	34.2±9.8	37	5:32	34.8±9.2
Yale University	48	8:40	29.4±8.2	22	3:19	31.8±6.9

[a] Alzheimer's Disease Neuroimaging Initiative - Department of Defense.
[b] Injury & Traumatic Stress Clinical Consortium.

models employed study site as a random factor to model random intercepts. The $LME_{INT+SLP}$ modeled both the site-specific random intercepts and age-related random slopes to reflect different age-related slopes in cortical thickness across sites. Bonferroni correction was employed for multiple testing of 148 cortical regions with a corrected $\alpha = 0.0003$ (0.05/148). The functions *lm()* and *lmer()* (from the R package *lme4) were* used to calculate the unstandardized regression coefficients for lineal models and the random effects models, respectively. The R package *lmerTest* was utilized to extract the statistical significance of models. The fitted curves in this manuscript were made using default settings (i.e. loess) of the R ggplot2 function *geom_smooth()*.

The number of regions with significant findings and the magnitude of effect size was compared separately between the 4 harmonization methods. A chi-squared test based on the function *chisq.test()* (from the native stats package in R) was used to compare the number of cortical regions showing significant effects. The region-specific regression coefficients were compared using repeated-measures ANOVA based on the function *aov_ez()* (from the R package *afex*). If the omnibus ANOVA results were statistically significant, then post-hoc pairwise comparisons of the 4 harmonization methods were conducted using the function *pairs()* (from the R package *emmeans*). The *p*-values were adjusted using the Bonferroni method for the 6 pairwise comparisons made with the outputs of the 4 harmonization methods.

3. Results

As shown in **Fig. 1** and the **interactive plot** at https://4n8ygg-delinsun.shinyapps.io/SDL_Shiny/, data distribution and age-related slops are largely modulated by site. Visual evidence of a non-linear age effect in participants under 20 years originate from 3 sites (Duke-De Bellis; Leiden University; University of Washington). Therefore, it is paramount and meaningful to harmonize the data by removing site effects.

3.1. Distribution of non-harmonized, ComBat harmonized, and ComBat-GAM harmonized data

Distributions of non-harmonized and harmonized data are shown in **Fig. 2**. Relative to non-harmonized data, ComBat (controls: range of *t-values*: [−10.120,−3.225], *p-values*: [<0.001,0.006] corrected; PTSD: *t-values*: [−9.653,−3.475], *p-values*: [<0.001,0.003] corrected; across regions) and ComBat-GAM harmonized data (controls: *t-values*: [−10.046,−1.856], *p-values*: [<0.001,0.207] corrected; PTSD: *t-values*: [−9.590,−2.284], *p-values*: [<0.001,0.078] corrected; across regions) resulted in smaller differences overall between the site-specific data and the mean across sites. There was no significant difference between ComBat and ComBat-GAM harmonized data (controls: *t-values*: [−2.373,0.075], *p-values*: [0.064,0.999] corrected; PTSD: *t-values*: [−2.183,0.066], *p-values*: [0.100,0.999] corrected; across regions).

There was no significant difference in the site-specific standard deviations between all data-pairings across all regions (controls: *t-values*: [−0.961,1.995], *p-values*: [0.154,0.999] corrected; PTSD: *t-values*: [−1.174,2.354], *p-values*: [0.066,0.999] corrected; across regions).

3.2. Main effect of age

As shown in **Fig. 3**, linear age-related trends are evident with ComBat harmonization, whereas non-linear trends are evident with ComBat-GAM harmonization with a dramatic decline in cortical thickness before 20, and a relatively slow decline after 20 years. This pattern holds true for both PTSD and control groups.

As shown in **Figs. 4A** and **5**, the number of regions showing a significant main effect of age was significantly different across harmonization methods ($X^2(3) = 89.658$, $p < 0.001$). The age-related declines in cortical thickness were detected by ComBat-GAM and ComBat in 147

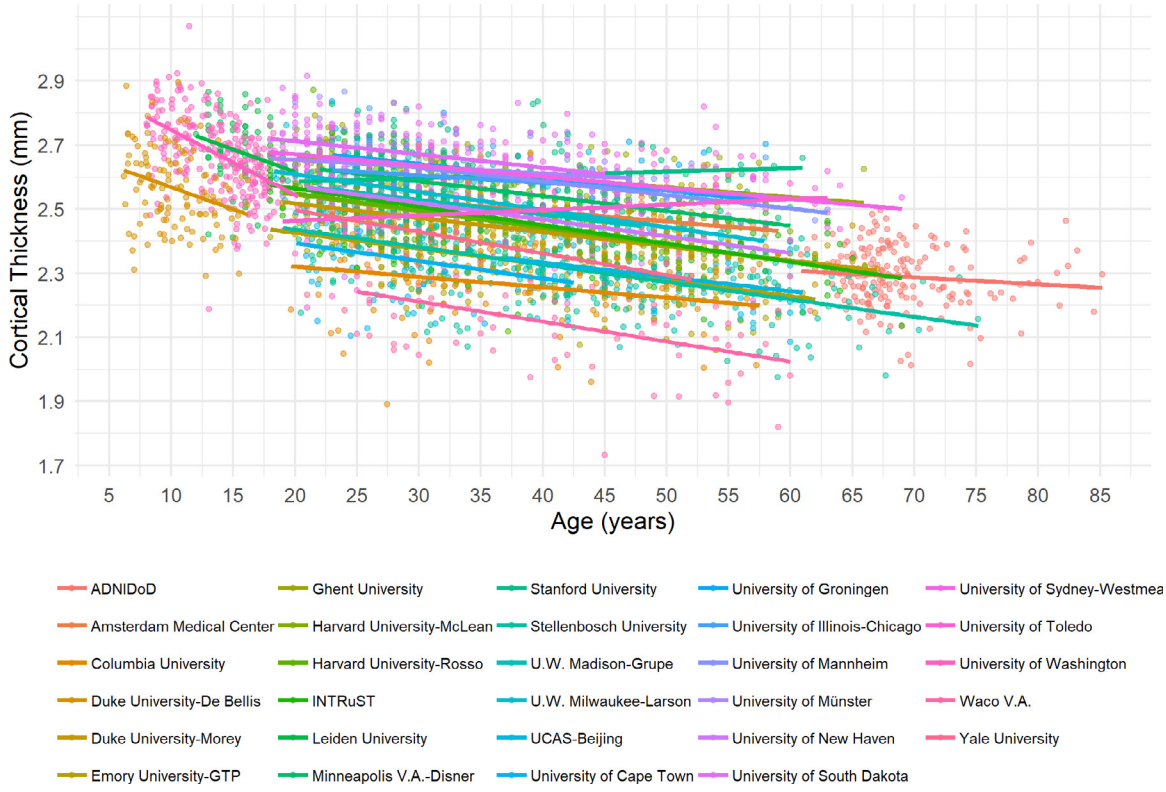

Fig. 1. Scatter plots of mean cortical thickness averaged across regions for each study site. Data distribution and age-related linear trends are markedly different across sites. Mean cortical thickness averaged across regions is shown to avoid regional biases. Participants are color-coded based on study site.

Table 2
Percent of 148 regions showing statistical significance.

Effects	Harmonization methods				Chi-squared test	
	LME_{INT}	$LME_{INT+SLP}$	ComBat	ComBat-GAM	*statistics*	*p (Bonferroni corr.)*
Age	98.0	76.4	99.3	99.3	89.658	<0.001
Diagnosis	1.4	1.4	3.4	20.9	63.704	<0.001
Age x Diagnosis	0	0	0	2.7	12.082	0.007
Sex	17.6	19.6	29.1	29.1	9.114	0.028
Sex x Diagnosis	0	0	0	0	NA	NA

Note: LME_{INT}, LME models site-specific random intercept. $LME_{INT+SLP}$, LME models both site-specific random intercepts and age-related random slopes.

(99.3%) regions, by LME_{INT} in 145 (98.0%) regions, and by $LME_{INT+SLP}$ in 113 (76.4%) regions, see Table 2. As shown in Table 3, the ratio of detection (> 95%) indicated that the significant regions detected by one method was also detected by another method except for $LME_{INT+SLP}$, which is less efficiently (< 80%) in replicating the findings of age effects detected by the other methods.

The regression coefficients were significantly different across harmonization methods ($F(1.6, 231.8) = 207.13$, $p < 0.001$). As shown in Fig. 4**B** and Table 4, ComBat-GAM produced stronger estimates of age-related declines in cortical thickness than the other methods, while the other three methods were not significantly different from each other.

3.3. Main effect of diagnosis

The number of regions showing a main effect of diagnosis was significantly different across harmonization approaches ($X^2(3) = 63.704$, $p < 0.001$). As shown in Fig. 6**A**, and Table 2, case-related reductions in cortical thickness were found by ComBat-GAM in 31 (20.9%) regions, by ComBat in 5 (3.4%) regions, by LME_{INT} and by $LME_{INT+SLP}$ in 2 (1.4%) regions. As shown in Fig. 7, the regions discovered by ComBat-GAM include those within the salience network (SN; bilateral insula regions), executive control network (ECN; bilateral intraparietal sulcus and supramarginal gyri), default mode network (DMN; left ventromedial prefrontal cortex, and bilateral precuneus), and bilateral superior and inferior temporal gyri and sulci, which are consistent with previous reports (Shalev et al., 2017). As shown in Table 3, the significant regions detected by LME_{INT} were also detected by ComBat, and the significant regions detected by $LME_{INT+SLP}$ were also detected by ComBat-GAM, while the opposite was not true (ratio of detection <= 40%).

Regression coefficients were different across harmonization methods ($F(1.4, 205.1) = 335.79$, $p < 0.001$). As shown in Fig. 6**B** and Table 4, ComBat-GAM produced stronger estimates of case-related cortical thickness reduction as well as weaker estimates of case-related cortical thickness increase than the other three methods, and ComBat produced stronger estimates of case-related cortical thickness reduction as well as weaker estimates of case-related cortical thickness increase than the two LME methods.

3.4. Age by diagnosis interaction

As shown in Fig. 8**A**, significant age by diagnosis interactions were detected by ComBat-GAM in 4 (2.7%) regions, while no significant in-

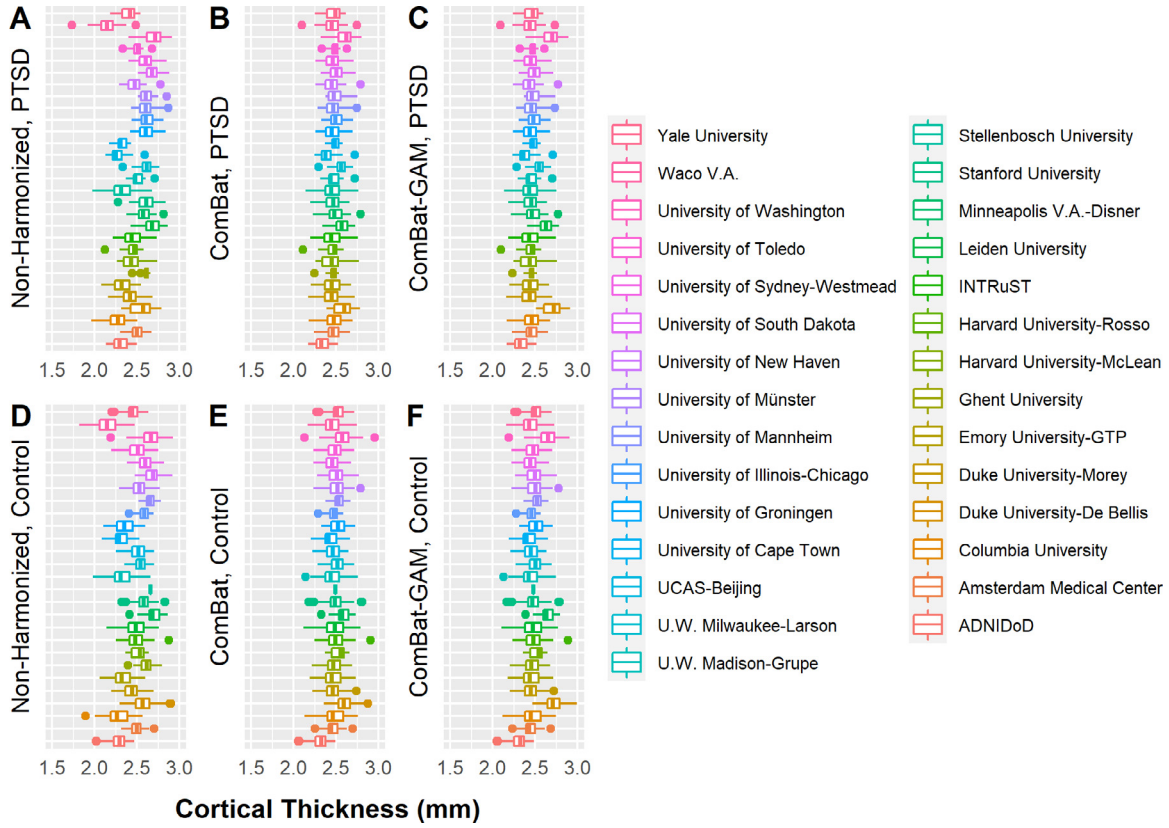

Fig. 2. Site-specific cortical thickness averaged across regions. Non-harmonized (A), ComBat harmonized (B), and ComBat-GAM harmonized (C) data in participants with PTSD. Non-harmonized (D), ComBat harmonized (E), and ComBat-GAM harmonized (F) data in trauma-exposed controls. The order of sites in the figure is consistent with the order of site names in the legend from top to bottom to facilitate with interpretation. Compared to non-harmonized data, ComBat and ComBat-GAM lead to smaller differences between site-specific data and the mean values averaged across sites, and they do not change the site-specific standard deviations for cortical thickness. The effects of harmonization by LME models cannot be shown here because data harmonization and statistical analyses are inseparable in LME methods. Mean cortical thickness averaged across regions is shown to minimize regional biases. The boxplots were made using the default settings of the R ggplot2 function *geom_boxplot()*. The lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles). The upper whisker extends from the hinge to the largest value no further than 1.5 * IQR from the hinge (where IQR is the inter-quartile range, or distance between the first and third quartiles). The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. Data beyond the end of the whiskers are called "outlying" points and are plotted individually.

teractions were detected by ComBat, LME_{INT}, and $LME_{INT+SLP}$. ComBat-GAM outperformed the other methods in detecting this interaction effect ($X^2(3) = 12.082$, $p = 0.007$), see Table 2. As shown in Fig. 9, age-related declines in cortical thickness were slower in cases than controls for 4 regions: right posterior-dorsal part of the cingulate gyrus, right marginal branch of the cingulate sulcus, right inferior temporal gyrus, and right fusiform gyrus. The linear (**Fig. S1**) and non-linear (**Fig. S2**) fits of the age-related distributions of cortical thickness harmonized by ComBat-GAM in these regions are shown in the supplementary materials.

Regression coefficients differed across harmonization methods ($F(1.3, 197.3) = 246.41$, $p < 0.001$). As shown in Fig. 8**B** and Table 4, ComBat-GAM compared to the other methods produced weaker estimates of age-related declines in cortical thickness in cases than controls, and both ComBat and LME_{INT} compared to $LME_{INT+SLP}$ produced weaker estimates of age-related declines in cortical thickness in cases than controls.

3.5. Main effect of sex

The number of regions showing a significant main effect of sex was significantly different across harmonization methods ($X^2(3) = 9.114$, $p = 0.028$). As shown in Fig. 10**A** and Table 2, the differences between males and females in cortical thickness were detected by ComBat-GAM and by ComBat in 43 (29.1%) regions, by LME_{INT} in 26 (17.6%) regions, and by $LME_{INT+SLP}$ in 29 (19.6%) regions. As shown in Fig. 11, The analyses based on ComBat-GAM harmonization showed that females had greater cortical thickness than males in bilateral precentral and postcentral regions, bilateral middle cingulate cortex, bilateral superior frontal gyri, bilateral angular gyri, bilateral medial occipito-temporal sulci and lingual sulci, left frontal pole, left superior temporal sulci, and right parahippocampal gyrus. By contrast, males had greater cortical thickness than females in bilateral inferior temporal regions, left rectus, left planum polare of the superior temporal gyrus, left vertical ramus of the anterior segment of the lateral sulcus, bilateral calcarine sulci, left insula, left inferior and middle frontal sulci, left orbital sulci, right ventral posterior cingulate cortex, right temporal pole. As shown in Table 3, most regions showing statistical significance detected by the LME methods were also detected by ComBat and ComBat-GAM (ratio of detection > 90%), and the opposite is not true (ratio of detection <= 70%).

Regression coefficients were different across harmonization methods ($F(1.8, 259.6) = 123.25$, $p < 0.001$). As shown in Fig. 10**B** and Table 4, ComBat-GAM compared to the other methods produced stronger estimates of cortical thickness reduction in females than males as well as weaker estimates of cortical thickness increase in females than males. ComBat compared to LME methods as well as LME_{INT} compared to $LME_{INT+SLP}$ produced stronger estimates of cortical thickness reduction in females than males as well as weaker estimates of cortical thickness increase in females than males.

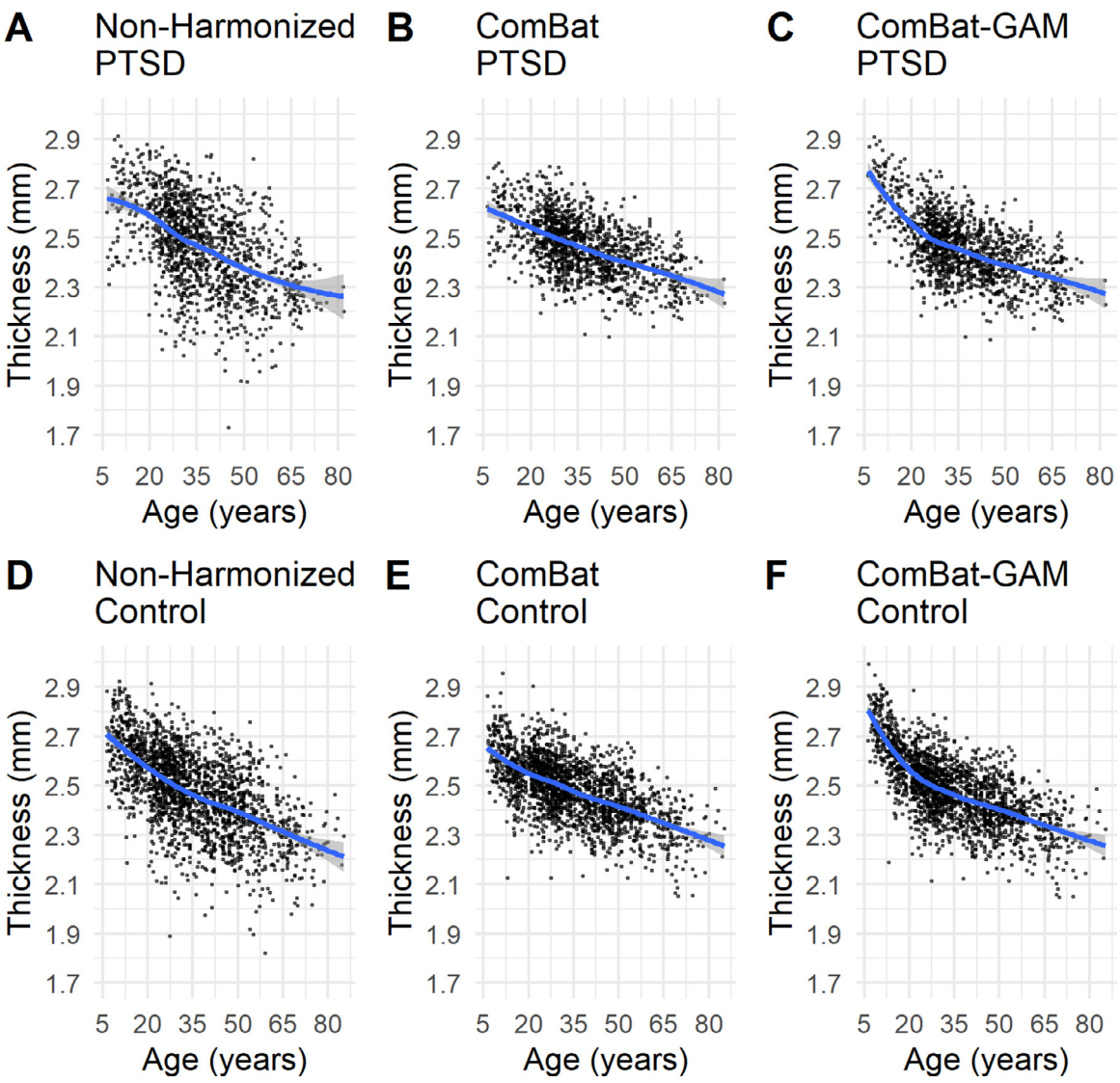

Fig. 3. Scatter plots and non-linear trends of mean cortical thickness averaged across regions. Non-harmonized (A), ComBat harmonized (B), and ComBat-GAM harmonized (C) data in participants with PTSD. Non-harmonized (D), ComBat harmonized (E), and ComBat-GAM harmonized (F) data in controls. Both ComBat and ComBat-GAM reduce variances. ComBat-GAM is superior to ComBat at capturing the age-related non-linear trends in cortical thickness. Mean cortical thickness averaged across regions is shown to avoid biases by particular regions. The fit curves were made based on the default settings (i.e. *loess*) of the R ggplot2 function *geom_smooth()*. The shaded regions represent the 95% confidence intervals.

3.6. Sex by diagnosis interaction

As shown in Fig. 12**A**, no significant sex by diagnosis interactions were found using data from any of the four methods. Regression coefficients were significantly different across harmonization approaches ($F(1.2, 178.3) = 12.40$, $p < 0.001$). As shown in Fig. 12**B** and Table 4, ComBat-GAM compared to the other methods produced stronger estimates of cortical thickness reduction in females relative to males in cases than controls, as well as weaker estimates of cortical thickness increase in females relative to males in cases than controls. ComBat compared to the $LME_{INT+SLP}$ methods produced stronger estimates of cortical thickness increase in females compared to males in cases than controls, as well as weaker estimates of cortical thickness reduction in females compared to males in cases than controls.

3.7. Results after removing sites with children, adolescents, and older participants

To test whether our findings were influenced by the data from children, adolescents, and very old participants, we re-analyzed the data after removing 3 sites with participants under 20 years and one site with older participants (~70 years). ComBat-GAM, ComBat, and LME_{INT} detected more regions with age-related cortical thinning compared to $LME_{INT+SLP}$. Both ComBat and ComBat-GAM compared to two LME methods detected more regions with sex-related cortical thickness differences. There was no significant difference among harmonization methods in detecting other effects. More details see **Supplementary results** section, and **Table S4, S5, and S6**.

4. Discussion

We compared the performance of four harmonization methods by applying them to cortical thickness data in participants grouped into clinical cases and controls from 29 different sites. The four harmonization methods included LME_{INT}, $LME_{INT+SLP}$, ComBat, and ComBat-GAM. We acknowledge that the number of regions reaching significance by any method does not necessarily reflect the *ground truth*, but the principle goal of harmonization is to convert uncalibrated measurements from multiple sources to be more comparable to each other. As summarized in Table 2, ComBat-GAM, ComBat, and LME_{INT} detected more regions with

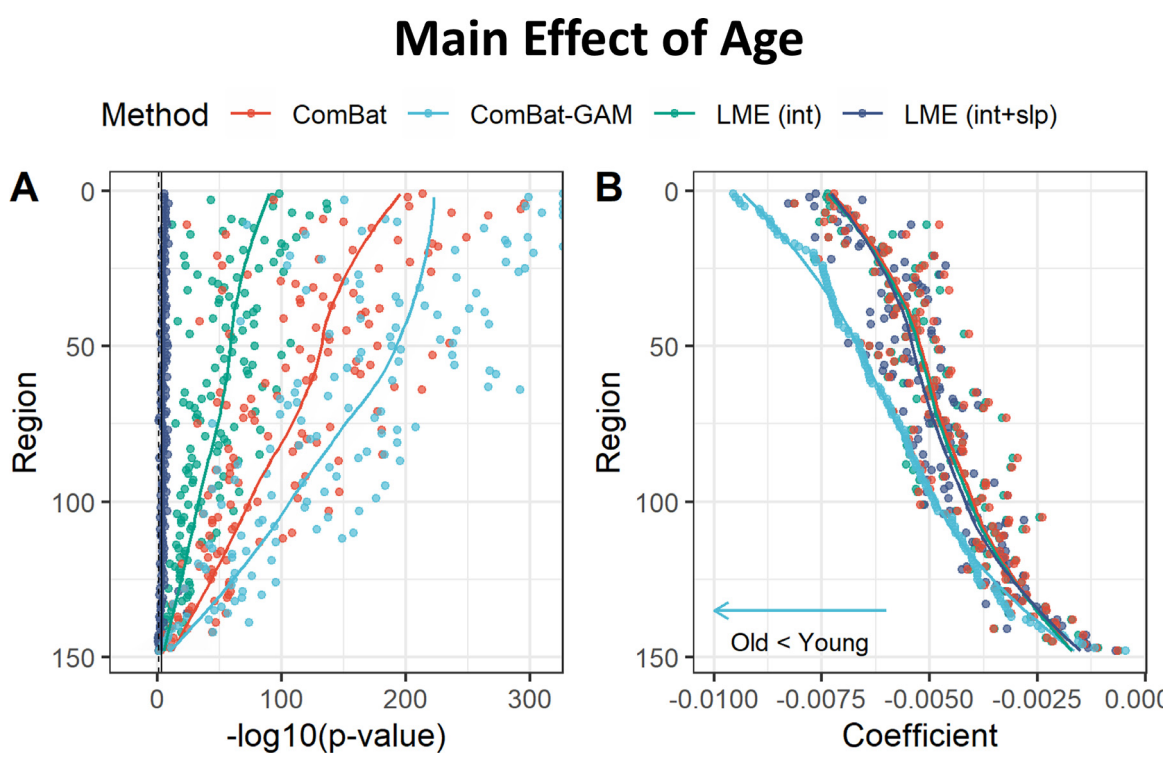

Fig. 4. Main effect of age. (A) Negative log-transformed statistical significance, i.e. $-\log_{10}(p)$. All four methods can detect multiple regions showing significance. The dashed and solid vertical lines represent thresholds $p = 0.05$ (uncorrected) and $p = 0.05$ (Bonferroni corrected), respectively. (B) Magnitude of regression coefficients. ComBat-GAM compared to the other methods provided stronger estimation of age-related cortical thickness reduction. The ordering of regions from top to bottom in both (A) and (B) is by ascending order of regression coefficients from cortical thickness data harmonized by ComBat-GAM. LME_{INT}, LME models site-specific random intercept. $LME_{INT+SLP}$, LME models both site-specific random intercepts and age-related random slopes. The fit curves were made based on the default settings (i.e. *loess*) of the R ggplot2 function *geom_smooth()*.

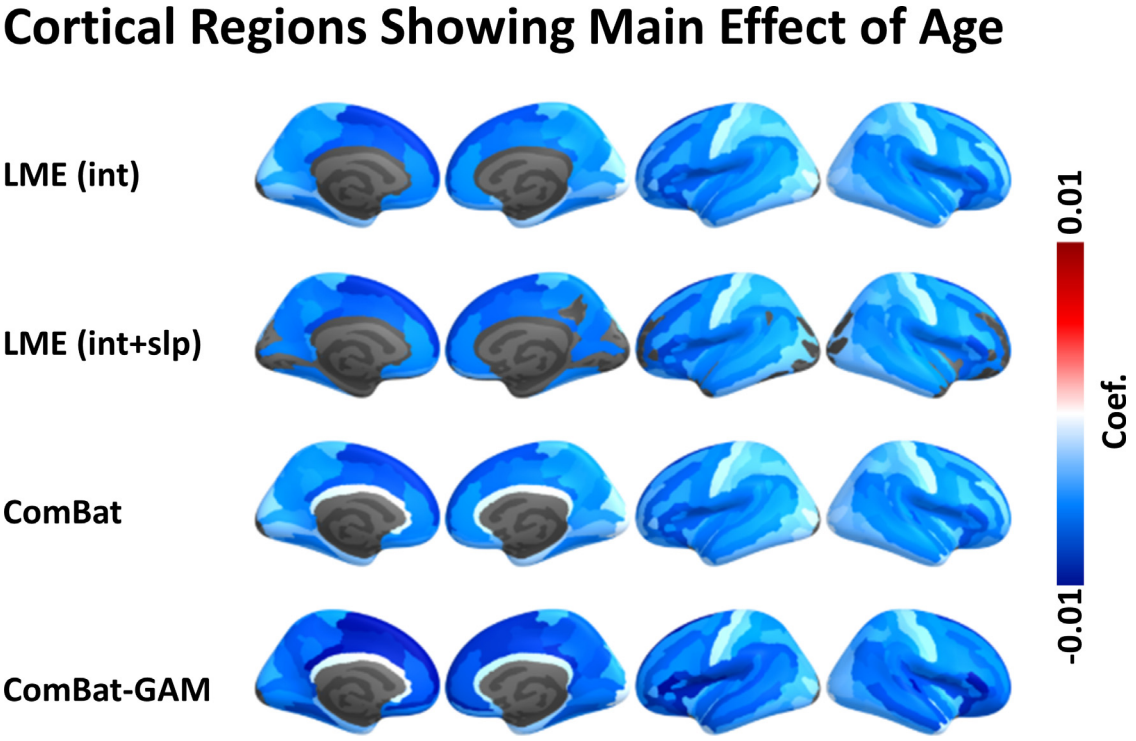

Fig. 5. Regions with significant main effect of age. The color bar represents the magnitude of the regression coefficient. $LME_{INT+SLP}$ compared to the other methods detected fewer regions showing significant age effect. Cooler colors represent stronger age-related declines in cortical thickness. LME_{INT}, LME models site-specific random intercept. $LME_{INT+SLP}$, LME models both site-specific random intercepts and age-related random slopes.

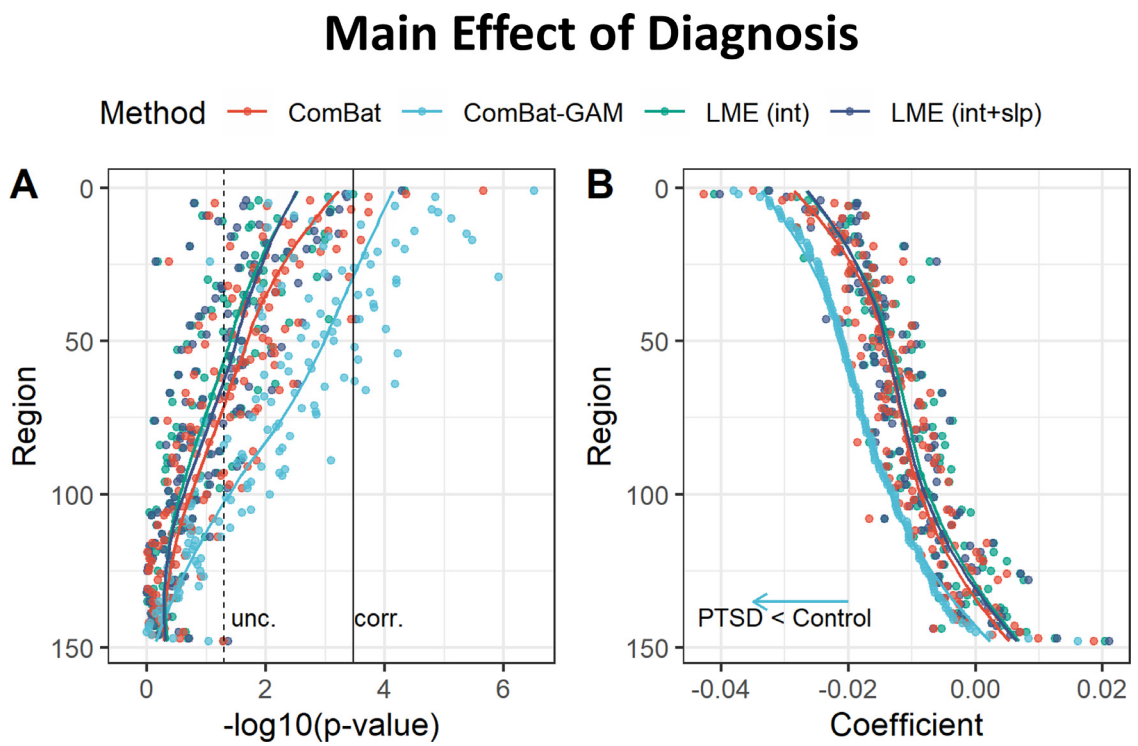

Fig. 6. Main effect of diagnosis. (A) Negative log-transformed statistical significance, i.e. $-\log_{10}(p)$. ComBat-GAM compared to the other methods detected more regions showing statistical significance. The dashed and solid vertical lines represent thresholds $p = 0.05$ (uncorrected) and $p = 0.05$ (Bonferroni corrected), respectively. (B) Magnitude of regression coefficients. ComBat-GAM compared to the other methods provided stronger estimation of case-related cortical thickness reduction as well as weaker estimation of case-related cortical thickness increase. The ordering of regions from top to bottom in both (A) and (B) is by ascending order of regression coefficients from cortical thickness data harmonized by ComBat-GAM. LME_{INT}, LME models site-specific random intercept. $LME_{INT+SLP}$, LME models both site-specific random intercepts and age-related random slopes. The fit curves were made based on the default settings (i.e. *loess*) of the R ggplot2 function *geom_smooth()*.

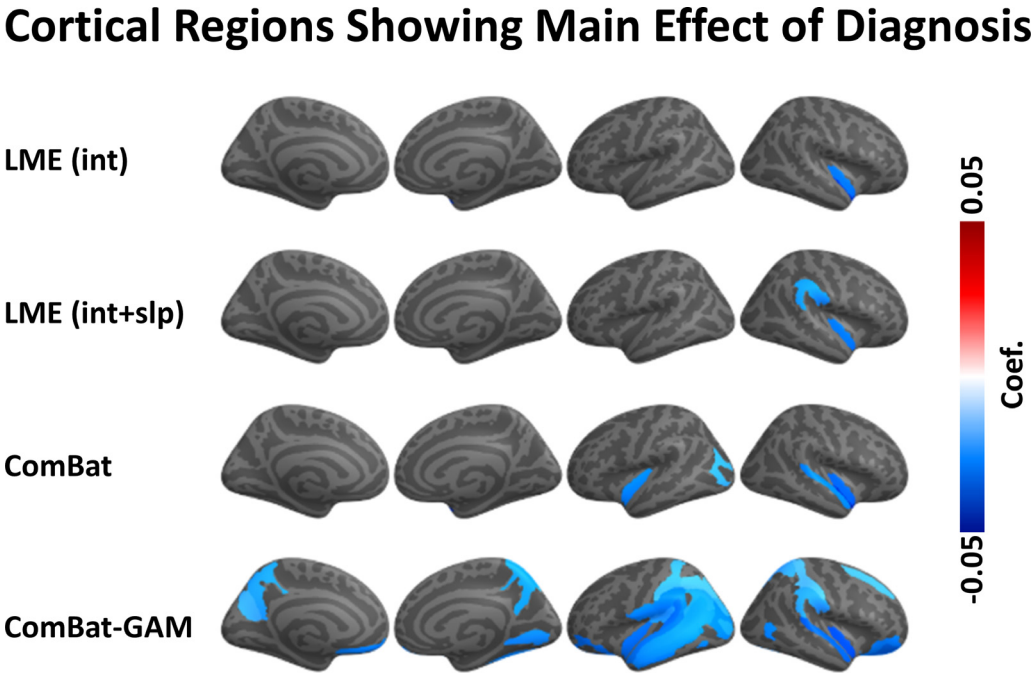

Fig. 7. Regions with a significant main effect of diagnosis. ComBat-GAM compared to the other methods detected more regions showing significant case-control difference. The color bar represents the magnitude of the regression coefficient. Cooler colors mean lower cortical thickness in PTSD than controls. LME_{INT}, LME models site-specific random intercept. $LME_{INT+SLP}$, LME models both site-specific random intercepts and age-related random slopes.

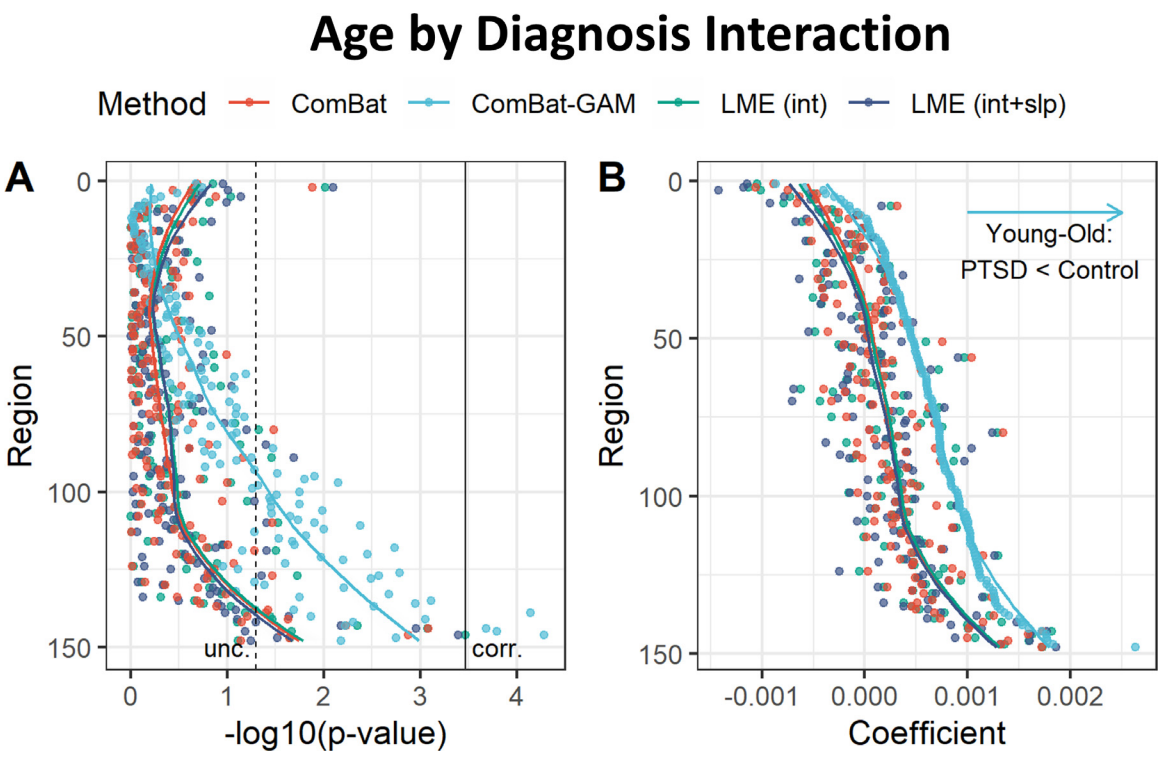

Fig. 8. Interaction of age and diagnosis. (A) Negative log-transformed statistical significance, i.e. $-\log_{10}(p)$. Only ComBat-GAM detected four regions showing statistical significance after correction. The dashed and solid vertical lines represent thresholds $p = 0.05$ (uncorrected) and $p = 0.05$ (Bonferroni corrected), respectively. (B) Magnitude of regression coefficients. ComBat-GAM compared to the other methods produced weaker estimates of age-related declines in cortical thickness in cases than controls. The ordering of regions from top to bottom in both (A) and (B) is by ascending order of regression coefficients from cortical thickness data harmonized by ComBat-GAM. LME_{INT}, LME models site-specific random intercept. $LME_{INT+SLP}$, LME models both site-specific random intercepts and age-related random slopes. The fit curves were made based on the default settings (i.e. *loess*) of the R ggplot2 function *geom_smooth()*.

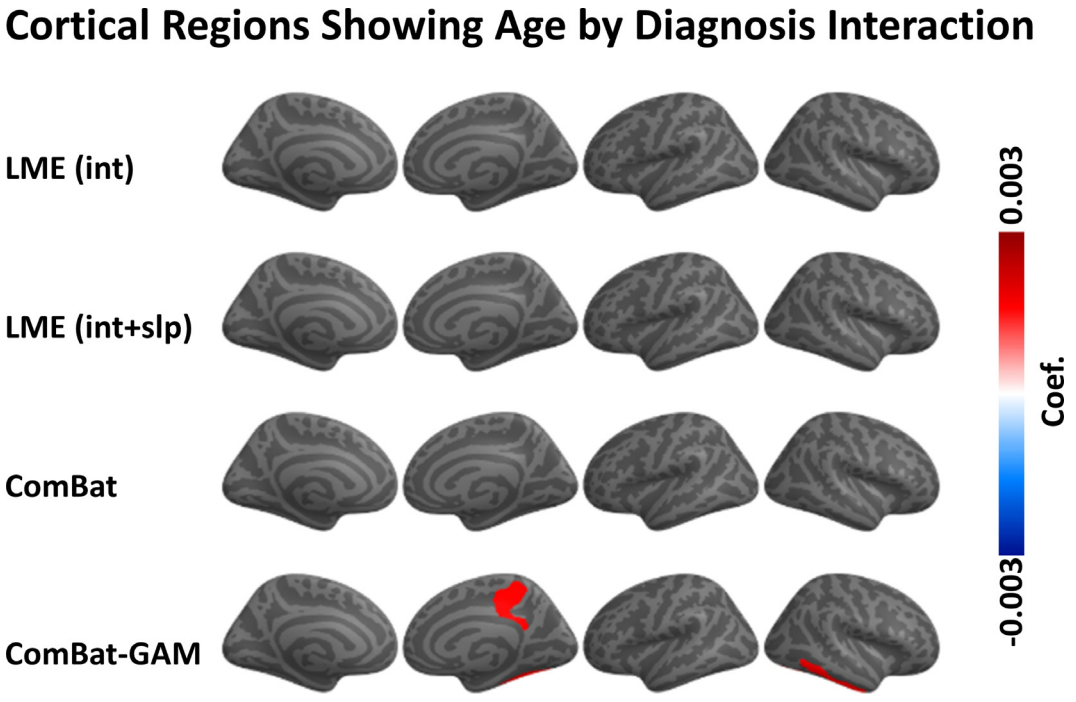

Fig. 9. Regions show significant age by diagnosis interaction. Only ComBat-GAM detected four regions showing statistical significance. The color bar represents the magnitude of the regression coefficient. Warmer colors mean that age-related declines in cortical thickness are smaller in PTSD than controls.

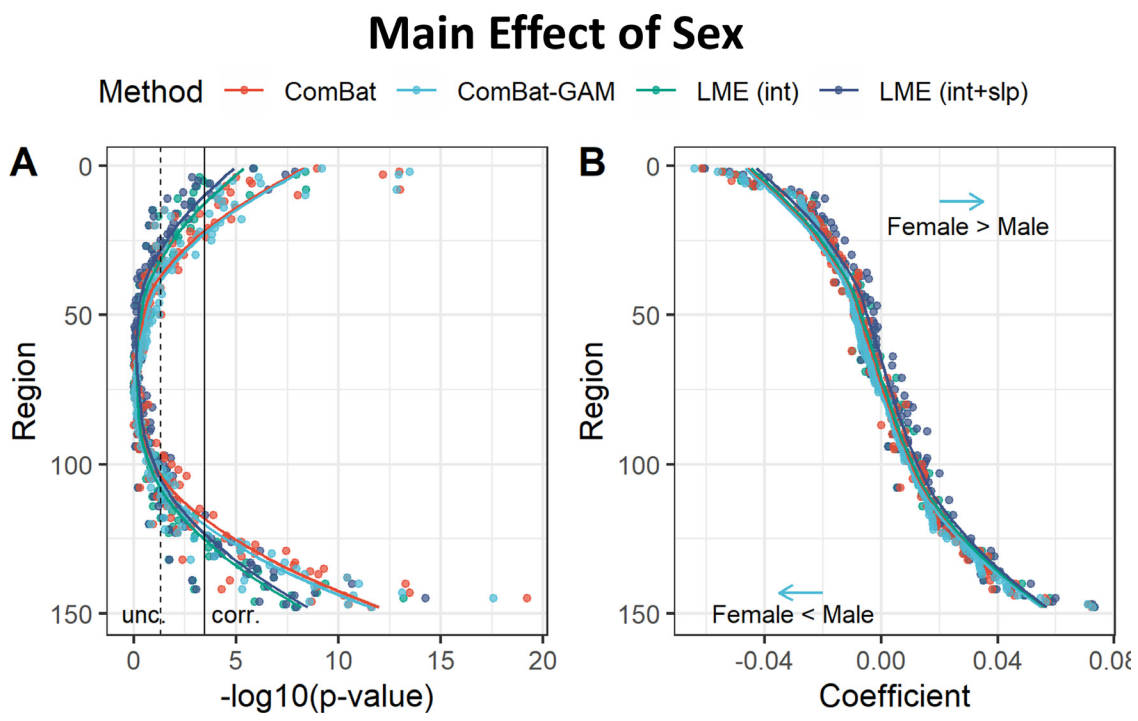

Fig. 10. Main effect of sex. (A) Negative log-transformed statistical significance, i.e. $-\log_{10}(p)$. The dashed and solid vertical lines represent thresholds $p = 0.05$ (uncorrected) and $p = 0.05$ (Bonferroni corrected), respectively. (B) Magnitude of regression coefficients. ComBat-GAM compared to the other methods produced stronger estimates of cortical thickness reduction in females than males as well as weaker estimates of cortical thickness increase in females than males. The ordering of regions from top to bottom in both (A) and (C) is by ascending order of regression coefficients from cortical thickness data harmonized by ComBat-GAM. LME_{INT}, LME models site-specific random intercept. $LME_{INT+SLP}$, LME models both site-specific random intercepts and age-related random slopes. The fit curves were made based on the default settings (i.e. *loess*) of the R ggplot2 function *geom_smooth()*.

Table 3
Ratio of detection (%) based on number of regions that met Bonferroni-corrected significance.

	LME_{INT}	$LME_{INT+SLP}$	ComBat	ComBat-GAM
	Effect of Age			
LME_{INT}	–	77.9	100	100
$LME_{INT+SLP}$	100	–	100	100
ComBat	98.6	76.9	–	100
ComBat-GAM	98.6	76.9	100	–
	Effect of Diagnosis			
LME_{INT}	–	50	100	50
$LME_{INT+SLP}$	50	–	50	100
ComBat	40	20	–	80
ComBat-GAM	3.2	6.5	12.9	–
	Age by Diagnosis Interaction			
LME_{INT}	–	NA	NA	NA
$LME_{INT+SLP}$	NA	–	NA	NA
ComBat	NA	NA	–	NA
ComBat-GAM	0	0	0	–
	Effect of Sex			
LME_{INT}	–	100	100	100
$LME_{INT+SLP}$	89.7	–	100	93.1
ComBat	60.5	67.4	–	93
ComBat-GAM	60.5	62.8	93	–
	Sex by Diagnosis Interaction			
LME_{INT}	–	NA	NA	NA
$LME_{INT+SLP}$	NA	–	NA	NA
ComBat	NA	NA	–	NA
ComBat-GAM	NA	NA	NA	–

Note: The *ratio of detection* is defined as the proportion of cortical regions showing statistical significance that were identified by the methods in rows were also detected by the methods in columns. Higher ratio of detection means that the method in columns was as effective as the method in rows at detecting significance. NA indicates not available because no significant finding was detected by the method in rows.

age-related cortical thinning compared to $LME_{INT+SLP}$ (Figs. 4**A** and 5). Consistent with our *a priori* hypothesis, ComBat-GAM harmonization uncovered more regions with significant case-related reductions in cortical thickness (Figs. 6**A** and 7), and more regions displaying slower rates of age-related cortical thinning in cases than controls compared to the other methods (Figs. 8**A** and 9). ComBat and ComBat-GAM outperformed LME methods in detecting sex-related differences (Figs. 10**A** and 11), but not sex by diagnosis interactions (Fig. 12**A**). As summarized in Table 3, most regions showing significant effects of age and sex detected by LME methods were also detected by ComBat and ComBat-GAM, while the opposite was not true, except that LME_{INT} performed comparably to ComBat and ComBat-GAM for the main effect of age. Regression coefficients (Table 4) showed that compared to other methods, ComBat-GAM produced stronger estimates of age-related declines in cortical thickness (Fig. 4**B**), stronger estimates of case-related cortical thickness reduction (Fig. 6**B**), weaker estimates of age-related declines in cortical thickness in cases than controls (Fig. 8**B**), stronger estimates of cortical thickness reduction in females than males as well as weaker estimates of cortical thickness increase in females than males (Fig. 10**B**), stronger estimates of cortical thickness reduction in females relative to males in cases than in controls, and weaker estimates of cortical thickness increase in females relative to males in cases than in controls (Fig. 12**B**).

ComBat models the expected values of the imaging features as a linear combination of the biological variables and the site effects whose error term is modulated by additional site-specific scaling factors (Fortin et al., 2018). It also uses empirical Bayes to improve the estimation of the model parameters in studies with small sample size. Radua et al. (2020) used cortical thickness, surface area, and subcortical volume data in cases and controls from ENIGMA-Schizophrenia to compare ComBat to random-effects meta-analysis and random-effects mega-analysis, which we term LME_{INT} in the present study. They reported that ComBat delivered more results that were statistically significant than

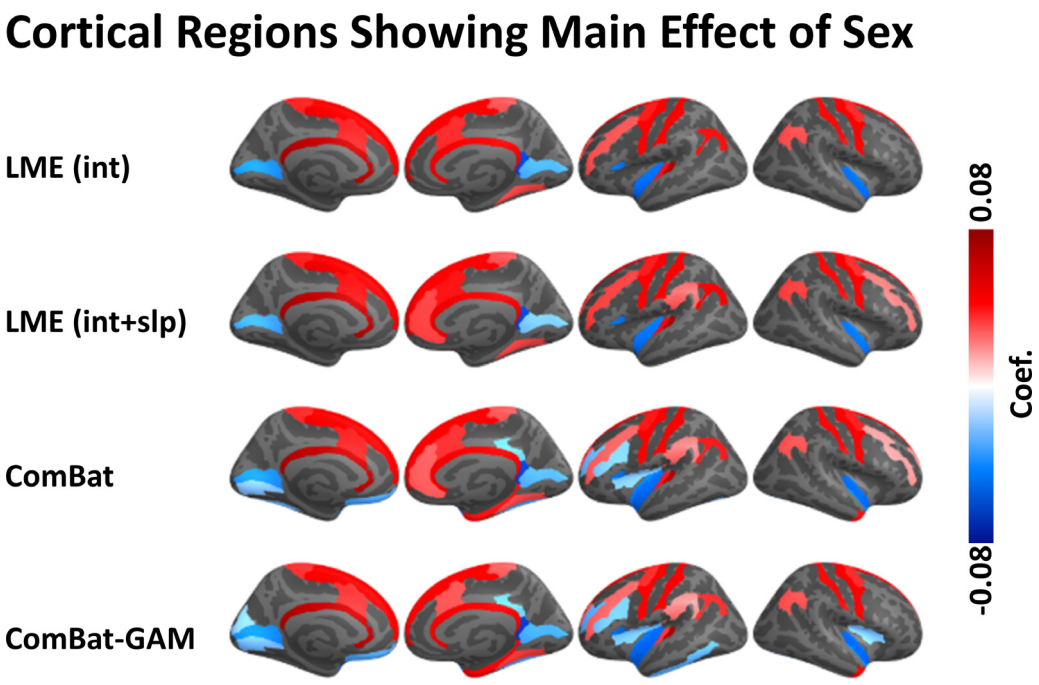

Fig. 11. Regions with a significant main effect of sex. The color bar represents the magnitude of the regression coefficient. Cooler (warmer) colors indicate lower (higher) cortical thickness in females compared to males.

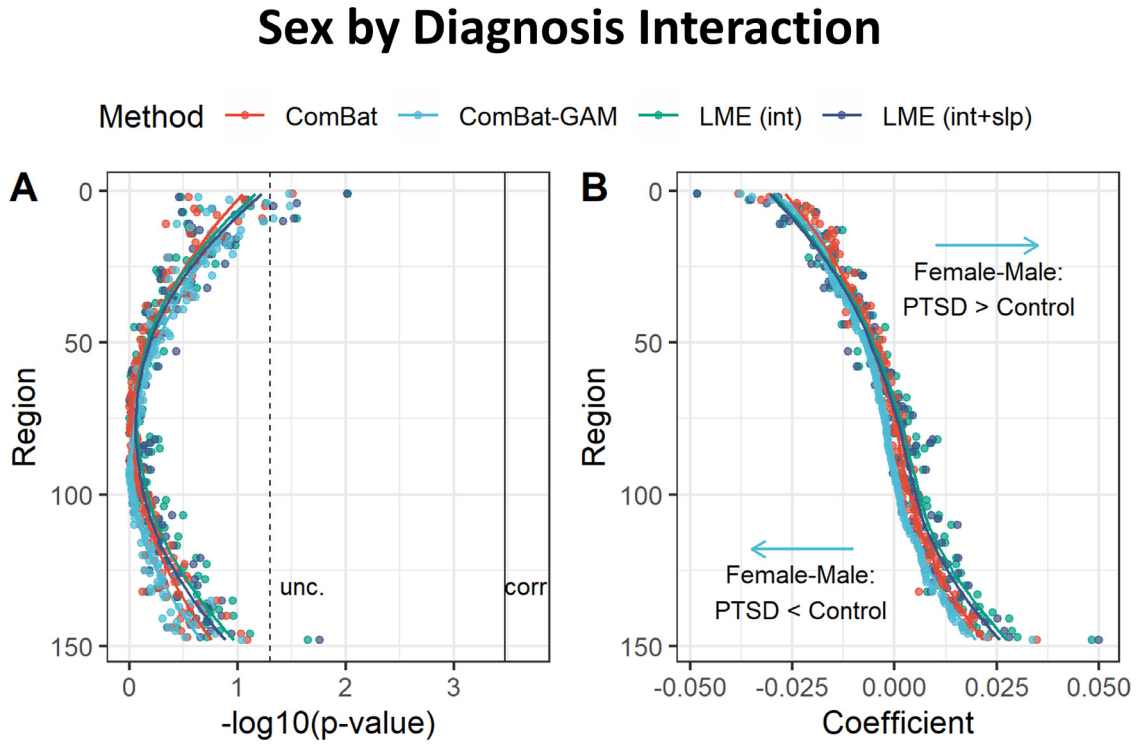

Fig. 12. Sex by diagnosis interaction. (A) Negative log-transformed statistical significance, i.e. $-\log_{10}(p)$. The dashed and solid vertical lines represent thresholds $p = 0.05$ (uncorrected) and $p = 0.05$ (Bonferroni corrected), respectively. None of the four methods detected significant regions. (B) The magnitude of regression coefficients. ComBat-GAM compared to the other methods produced stronger estimates of cortical thickness reduction in females relative to males in cases than controls, as well as weaker estimates of cortical thickness increase in females relative to males in cases than controls. The ordering of regions from top to bottom in both (A) and (B) is by ascending order of regression coefficients from cortical thickness data harmonized ComBat-GAM. LME_{INT}, LME models site-specific random intercept. $LME_{INT+SLP}$, LME models both site-specific random intercepts and age-related random slopes. The fit curves were made based on the default settings (i.e. *loess*) of the R ggplot2 function *geom_smooth()*.

Table 4
Comparisons of regression coefficients.

	LME_{INT}	$LME_{INT+SLP}$	ComBat	ComBat-GAM
	Effect of Age			
LME_{INT}	–	0.6e-04	−0.6e-04	10.8e-04
$LME_{INT+SLP}$	0.703	–	−1.2e-04	10.2e-04
ComBat	0.677	0.125	–	11.4e-04
ComBat-GAM	<0.001	<0.001	<0.001	–
	Effect of Diagnosis			
LME_{INT}	–	0.4e-03	1.2e-03	6.6e-03
$LME_{INT+SLP}$	0.277	–	0.7e-03	6.2e-03
ComBat	<0.001	0.011	–	5.4e-03
ComBat-GAM	<0.001	<0.001	<0.001	–
	Age by Diagnosis Interaction			
LME_{INT}	–	3.8e-05	−0.3e-05	−44.4e-05
$LME_{INT+SLP}$	0.262	–	−4.1e-05	−48.2e-05
ComBat	0.999	0.197	–	−44.1e-05
ComBat-GAM	<0.001	<0.001	<0.001	–
	Effect of Sex			
LME_{INT}	–	−1.8e-03	0.6e-03	1.6e-03
$LME_{INT+SLP}$	<0.001	–	2.4e-03	3.4e-03
ComBat	0.012	<0.001	–	1.1e-03
ComBat-GAM	<0.001	<0.001	<0.001	–
	Sex by Diagnosis Interaction			
LME_{INT}	–	6.0e-04	−0.4e-04	12.0e-04
$LME_{INT+SLP}$	0.056	–	−6.3e-04	6.1e-04
ComBat	0.999	0.037	–	12.4e-04
ComBat-GAM	<0.001	0.049	<0.001	–

Note: The upper triangle of the matrix are the differences between regression coefficients from methods in rows and columns. Higher values mean that the method in columns lead to more negative (i.e., weaker positive coefficients, or stronger negative coefficients) estimates than the method in rows. The lower triangle represents the corresponding *p*-values (Bonferroni corrected).

random-effects meta-analyses, and slightly more than LME_{INT}. However, they did not report results of non-linear age effects on cortical thickness, which are well documented (Frangou et al., 2022; Pomponio et al., 2020; Walhovd et al., 2017), nor did they report on effects of group membership on age-related changes in cortical thickness. By contrast, Pomponio et al. (2020) developed ComBat-GAM to support harmonization of neuroimaging data with non-linearities related to age or other variables by investigating cortical and subcortical gray matter volumes in 10,477 healthy subjects ranging in age from 3 to 96 years collected at 18 sites. They concluded that ComBat-GAM is superior to ComBat at predicting age based on regional volume data. However, Pomponio et al. (2020) only investigated healthy participants, which lacked guidance on harmonization of data for case-control comparisons. Finally, prior studies did not report the magnitude of regression coefficients obtained from various harmonization methods, in spite of an urgent plea by researchers to understand how harmonization influences the output of statistical models run on harmonized data.

Our study sought to fill these gaps by formally comparing regression coefficients and the number of regions showing statistically significant results, including case-control differences in cortical thickness across the lifespan. As shown in **Fig. 2**, ComBat and ComBat-GAM led to smaller differences between site-specific data and the mean values averaged across sites, and they did not change the site-specific standard deviations for cortical thickness. These results demonstrated that both ComBat and ComBat-GAM are effective at minimizing the effects of site without distorting the data distribution. Harmonization with ComBat-GAM was the most effective at detecting case-control differences as evidenced by significantly more regional findings as compared to other harmonization methods. ComBat-GAM was one of the most effective methods at detecting age-effects in cortical thickness, and the only method to uncover regions with different rates of age-related cortical thinning in cases compared to controls. Furthermore, most of the regions showing statistical significance following harmonization with other methods were also detected following ComBat-GAM harmonization. Whereas we have no collateral information to corroborate the findings from ComBat-GAM harmonization pertaining to case-control differences or age-dependent case-control differences, the literature offers consistent evidence of age-related patterns of cortical thickness across the lifespan (Frangou et al., 2022; Mutlu et al., 2013). One caveat is that motion related artifact, which is associated with lower cortical thickness measurements, increases with age (Savalia et al., 2017). Consequently, reduced cortical thickness with aging may be partially artifactual. Nonetheless, **Fig. 3** shows concrete evidence of erroneous harmonization by ComBat that is handled correctly by ComBat-GAM. Our finding is corroborated by independent studies, which demonstrate that the highest cortical thickness occurs in childhood and that age is negatively correlated to cortical thickness with a steeper decline up to the third decade of life more gradually thereafter (Frangou et al., 2022; Mutlu et al., 2013). By contrast, ComBat harmonized the data along a linear pattern with age throughout the lifespan. Thus, ComBat-GAM harmonization may be advantageous, particularly for consortia studies with participants of all ages, particularly youth and young adults.

The performance of ComBat-GAM is attributable to its algorithm. LME models assume that the error terms follow the same normal distribution at all sites, which is rarely the case (Radua et al., 2020). ComBat overcomes this shortcoming by assuming different normal distributions at different sites for the error terms (Radua et al., 2020). ComBat-GAM further improves on ComBat by using a normal distribution as the prior for the intercept and an inverse-gamma distribution as the prior for the scale effect of the sites. It also uses generalized additive model (GAM) to capture non-linear variations in age-related changes in cortical thickness while avoiding overfitting (Pomponio et al., 2020).

In our study, participants at most sites were aged 20–60 years old, while volunteers from three sites were mostly below 20 years old, and participants from one site were mostly over 70 years old. We found that after removing the data from the four sites with either very young or very old participants, ComBat-GAM is not better than other harmonization methods at detecting regions with significant case-control differences and age by diagnosis interactions (see **supplementary results** section). We could not exclude the possibility that the superiority of ComBat-GAM versus the other methods is driven by overfitting data from sites with very young or very old participants. **Fig. 1** shows the data distributions of the four sites are consistent with the literature, with steeper cortical thickness declines in youth and flatter age-appropriate declines in older adults (Frangou et al., 2022; Mutlu et al., 2013). Furthermore, the three sites with participants < 20 years old exhibit similar slopes of age-related declines in cortical thickness. Therefore, rather than concluding that ComBat-GAM overfits data from children contributed by specific sites, there is stronger evidence to conclude that ComBat-GAM accurately captures nonlinear age trends in cortical thickness. Data from sites with a larger age range may address this concern more conclusively.

We found slower rates of age-related decline in cortical thickness in cases compared to controls for 4 regions, but only for data harmonized with ComBat-GAM (**Figs. 8** and **9**). As shown in supplementary **Figs. S1** and **S2**, cases exhibited lower cortical thickness compared to controls in youth and greater cortical thickness in elderly in the 4 regions. It is possible that PTSD induces more powerful cortical thinning in youth and delayed age-appropriate declines in cortical thickness in elderly. This explanation is partly consistent with previous findings that maltreated youth with versus without chronic PTSD have smaller volumes in the posterior brain structures (De Bellis et al., 2015). More studies are warranted to test whether case-control differences in age-related cortical thinning is overfit by ComBat-GAM.

A study by Ritchie et al. (2018) examined sex-differences in adults from UK Biobank (2750 females; males 2466; 44–77 years old) reported thicker cortex across most of the cortex in females than males except for the right insula. By contrast, harmonization with ComBat-GAM in our study showed that females have greater cortical thickness in prefrontal cortex, inferior parietal regions, and cingulate cortex, whereas males had greater cortical thickness in ventromedial prefrontal cortex, bilat-

eral insula, posterior cingulate areas, and occipital lobe (**Fig. 11**). This difference may be explained by the large difference of age range in the present study (6.2–85.2 years old) compared to Ritchie et al. (2018) (44–77 years old). As shown in **Fig. 3**, the slopes of age-related changes in cortical thickness are quite different between young (especially < 20 years old) and old participants. The significant differences between males and females in cortical thickness may be driven by the data of relatively young participants. We found that ComBat and ComBat-GAM outperformed LME harmonization methods for detecting differences between females and males. While we did not formally test harmonization methods to detect age-related sex differences in cortical thickness, Frangou et al., 2022 reported that age-related declines in mean cortical thickness were more rapid in males than females in the mid-life group (30–59 years), but not in the early-life group (3–29 years) and late-life group (61–90 years).

The comparison of the regression coefficients showed that the selection of harmonization methods may either overestimate or underestimate effects of interest, even though the corresponding comparisons of the number of regions exhibiting significant effects were identical between methods. These findings are critical to interpreting statistical outputs. For instance, the magnitude of reductions in cortical thickness per year are biased by the harmonization method being used.

In reporting that ComBat-GAM is more sensitive than other methods, we must be clear to specify our narrow definition of "sensitive", as the harmonization method that leads to the maximum number of brain regions with statistically significant effects. In fact, this metric does not necessarily determine better performance if we adopt a preferred definition, namely the method that produces results that are most consistent with the *ground truth*. Unfortunately, identifying ground truth is a challenging proposition, but we consider two options that may be informative and feasible. The first option is to acquire MRI scans and calculate cortical thickness from the same group of participants (or "travelling subjects") on a variety of scanner manufacturers and MRI facilities. However, a sufficient sample size is essential as it must contain (1) a representative number of cases and controls from (2) across the lifespan in (3) participants of both sexes, (4) scans at each MRI facility and on scanners from each manufacturer. This is required to avoid possible confounds from interactions of scanner type and age, scanner type and diagnosis, and scanner type and sex. A second option is to generate simulated data from a large enough sample of participants, sites, and MRI facilities. The simulated data could be generated by adding characteristic noise, covariance, and bias profiles for each scanner manufacturer and each MRI facility. The simulated data could then be harmonized with several tools of interest to determine the method that produces data that most closely resembles the pre-noised data. Along the same lines, the post-harmonization data and the pre-noised data could be modeled for case-control effects, age effects, and interaction effects. The results of statistical modeling on post-harmonization datasets could be compared to the results from modeling the pre-noised dataset. The harmonization method that leads to results that most closely resemble the results obtained from modeling the pre-noised data would be deemed most faithful to the ground truth. Scanning an appropriate phantom may add value to ascertaining the ground truth, but is unlikely to add value to characterizing the role of age, sex, and diagnosis on harmonization methods.

While our study focused on 4 widely adopted harmonization methods, these represent only a small number in a large array of available methods. There has been a recent explosion in methods that apply machine learning and other advanced multivariate techniques to tackle harmonization. More detailed discussions about machine learning in data harmonization please see supplementary section "Machine learning in data harmonization". The dawn of the big data age has heralded the need for harmonization methods that operate well beyond neuroimaging data to flexibly and extensibly harmonize manifold data types from social media, mobile devices, and sensors (Agarwal et al., 2013; Davatzikos, 2019). The rapid proliferation of data harmonization methods and the ubiquity of machine learning applications will require careful vetting and rigorous comparisons between competing methods using standard criteria for ascertaining harmonization performance. The urgent goal of advancing *open science* will be facilitated by developing and embracing advanced harmonization methods (Foster and Deardorff, 2017).

4.1. Limitations

There are four major limitations in the present study. Firstly, we investigated age-related changes in cortical thickness. However, only cross-sectional data was available. New approaches have been developed to harmonize data across scanners and sites as well as longitudinal visits (Beer et al., 2020; Dewey et al., 2019). Age-related cortical thinning estimated by one longitudinal study design was 3 times greater than cortical thinning from a cross-sectional study (Rast et al., 2018). Secondly, we only investigated cortical thickness, which is one of many brain measures that is disturbed in neuropsychiatric disorders. Further studies should investigate the performance of harmonization methods on multi-modal neuroimaging data with various anatomical, diffusion, functional, and clinical/behavioral measures. Thirdly, only three sites constituted participants under 20 years, and one site constituted participants over 70 years. After removing these data, ComBat-GAM did not outperform other harmonization methods in detecting regions with significant case-control differences and age by diagnosis interactions. Data from sites with a larger age range may address this concern more conclusively. Finally, we applied the same statistical model to the output of all harmonization methods to pinpoint differences between harmonization methods rather than statistical models or the interaction of harmonization methodology and statistical modeling. In the main text, the statistical model includes age, sex, and PTSD diagnosis as fixed factors. This model is simple and widely used in most psychiatric studies. We also consider age by diagnosis, and sex by diagnosis interactions because they are frequently tested in the literature. The statistical models listed in the main text may not fully reveal potential influences on cortical thickness, and the optimal statistical model may differ depending on the harmonization method. However, investigating potential interactions of harmonization method and statistical model are well beyond the scope of this study.

5. Conclusion

Cortical thickness data harmonized with ComBat-GAM relative to LME_{INT}, $LME_{INT+SLP}$, and ComBat is more sensitive at detecting significant case-control differences, and case-control differences that vary by age. Both ComBat and ComBat-GAM outperformed LME methods in detecting significant sex differences. ComBat-GAM provides stronger estimates of age-related declines in cortical thickness, stronger estimates of case-related cortical thickness reduction, weaker estimates of age-related declines in cortical thickness in cases than controls, stronger estimates of cortical thickness reduction in females than males, stronger estimates of cortical thickness reduction in females compared to males in cases than in controls. Our results support using ComBat-GAM to harmonize cortical thickness data across study sites to recover statistical power potentially lost by instrumental bias.

Conflicts of Interest

Dr. Abdallah has served as a consultant, speaker and/or on advisory boards for FSV7, Lundbeck, Psilocybin Labs, Genentech and Janssen, and editor of Chronic Stress for Sage Publications, Inc.; he has filed a patent for using mTOR inhibitors to augment the effects of antidepressants (filed on August 20, 2018). Dr. Davidson is the founder and president of, and serves on the board of directors for, the non-profit organization Healthy Minds Innovations, Inc. Dr. Jahanshad, Dr. Thompson and Dr. Ching received partial research support from Biogen, Inc. (Boston, USA) for research unrelated to the content of this manuscript. Dr. Krystal

is a consultant for AbbVie, Inc., Amgen, Astellas Pharma Global Development, Inc., AstraZeneca Pharmaceuticals, Biomedisyn Corporation, Bristol-Myers Squibb, Eli Lilly and Company, Euthymics Bioscience, Inc., Neurovance, Inc., FORUM Pharmaceuticals, Janssen Research & Development, Lundbeck Research USA, Novartis Pharma AG, Otsuka America Pharmaceutical, Inc., Sage Therapeutics, Inc., Sunovion Pharmaceuticals, Inc., and Takeda Industries; is on the Scientific Advisory Board for Lohocla Research Corporation, Mnemosyne Pharmaceuticals, Inc., Naurex, Inc., and Pfizer; is a stockholder in Biohaven Pharmaceuticals; holds stock options in Mnemosyne Pharmaceuticals, Inc.; holds patents for Dopamine and Noradrenergic Reuptake Inhibitors in Treatment of Schizophrenia, US Patent No. 5,447,948 (issued September 5, 1995), and Glutamate Modulating Agents in the Treatment of Mental Disorders, U.S. Patent No. 8,778,979 (issued July 15, 2014); and filed a patent for Intranasal Administration of Ketamine to Treat Depression. U.S. Application No. 14/197,767 (filed on March 5, 2014); US application or Patent Cooperation Treaty international application No. 14/306,382 (filed on June 17, 2014); Filed a patent for using mTOR inhibitors to augment the effects of antidepressants (filed on August 20, 2018). Dr. Schmahl is a consultant for Boehringer Ingelheim International GmbH. Dr. Stein has received research grants and/or consultancy honoraria from Lundbeck and Sun. Dr. Lebois reports unpaid membership on the Scientific Committee for the International Society for the Study of Trauma and Dissociation (ISSTD) and spousal license payment for Vanderbilt IP from Acadia Pharmaceuticals unrelated to the topic of this manuscript. All other authors have no conflicts of interest to declare.

Credit authorship contribution statement

Delin Sun: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing. **Gopalkumar Rakesh:** Data curation, Project administration, Funding acquisition. **Courtney C. Haswell:** Data curation, Project administration, Funding acquisition. **Mark Logue:** Conceptualization, Methodology, Writing – review & editing. **C. Lexi Baird:** Data curation, Project administration, Funding acquisition. **Erin N. O'Leary:** Data curation, Project administration, Funding acquisition. **Andrew S. Cotton:** Data curation, Project administration, Funding acquisition. **Hong Xie:** Data curation, Project administration, Funding acquisition. **Marijo Tamburrino:** Data curation, Project administration, Funding acquisition. **Tian Chen:** Data curation, Project administration, Funding acquisition. **Emily L. Dennis:** Data curation, Project administration, Funding acquisition. **Neda Jahanshad:** Data curation, Project administration, Funding acquisition. **Lauren E. Salminen:** Data curation, Project administration, Funding acquisition. **Sophia I. Thomopoulos:** Data curation, Project administration, Funding acquisition. **Faisal Rashid:** Data curation, Project administration, Funding acquisition. **Christopher R.K. Ching:** Data curation, Project administration, Funding acquisition. **Saskia B.J. Koch:** Data curation, Project administration, Funding acquisition. **Jessie L. Frijling:** Data curation, Project administration, Funding acquisition. **Laura Nawijn:** Data curation, Project administration, Funding acquisition. **Mirjam van Zuiden:** Data curation, Project administration, Funding acquisition. **Xi Zhu:** Data curation, Project administration, Funding acquisition. **Benjamin Suarez-Jimenez:** Data curation, Project administration, Funding acquisition. **Anika Sierk:** Data curation, Project administration, Funding acquisition. **Henrik Walter:** Data curation, Project administration, Funding acquisition. **Antje Manthey:** Data curation, Project administration, Funding acquisition. **Jennifer S. Stevens:** Data curation, Project administration, Funding acquisition. **Negar Fani:** Data curation, Project administration, Funding acquisition. **Sanne J.H. van Rooij:** Data curation, Project administration, Funding acquisition. **Murray Stein:** Data curation, Project administration, Funding acquisition. **Jessica Bomyea:** Data curation, Project administration, Funding acquisition. **Inga K. Koerte:** Data curation, Project administration, Funding acquisition. **Kyle Choi:** Data curation, Project administration, Funding acquisition. **Steven J.A. van der Werff:** Data curation, Project administration, Funding acquisition. **Robert R.J.M. Vermeiren:** Data curation, Project administration, Funding acquisition. **Julia Herzog:** Data curation, Project administration, Funding acquisition. **Lauren A.M. Lebois:** Data curation, Project administration, Funding acquisition. **Justin T. Baker:** Data curation, Project administration, Funding acquisition. **Elizabeth A. Olson:** Data curation, Project administration, Funding acquisition. **Thomas Straube:** Data curation, Project administration, Funding acquisition. **Mayuresh S. Korgaonkar:** Data curation, Project administration, Funding acquisition. **Elpiniki Andrew:** Data curation, Project administration, Funding acquisition. **Ye Zhu:** Data curation, Project administration, Funding acquisition. **Gen Li:** Data curation, Project administration, Funding acquisition. **Jonathan Ipser:** Data curation, Project administration, Funding acquisition. **Anna R. Hudson:** Data curation, Project administration, Funding acquisition. **Matthew Peverill:** Data curation, Project administration, Funding acquisition. **Kelly Sambrook:** Data curation, Project administration, Funding acquisition. **Evan Gordon:** Data curation, Project administration, Funding acquisition. **Lee Baugh:** Data curation, Project administration, Funding acquisition. **Gina Forster:** Data curation, Project administration, Funding acquisition. **Raluca M. Simons:** Data curation, Project administration, Funding acquisition. **Jeffrey S. Simons:** Data curation, Project administration, Funding acquisition. **Vincent Magnotta:** Data curation, Project administration, Funding acquisition. **Adi Maron-Katz:** Data curation, Project administration, Funding acquisition. **Stefan du Plessis:** Data curation, Project administration, Funding acquisition. **Seth G. Disner:** Data curation, Project administration, Funding acquisition. **Nicholas Davenport:** Data curation, Project administration, Funding acquisition. **Daniel W. Grupe:** Data curation, Project administration, Funding acquisition. **Jack B. Nitschke:** Data curation, Project administration, Funding acquisition. **Terri A. deRoon-Cassini:** Data curation, Project administration, Funding acquisition. **Jacklynn M. Fitzgerald:** Data curation, Project administration, Funding acquisition. **John H. Krystal:** Data curation, Project administration, Funding acquisition. **Ifat Levy:** Data curation, Project administration, Funding acquisition. **Miranda Olff:** Data curation, Project administration, Funding acquisition. **Dick J. Veltman:** Data curation, Project administration, Funding acquisition. **Li Wang:** Data curation, Project administration, Funding acquisition. **Yuval Neria:** Data curation, Project administration, Funding acquisition. **Michael D. De Bellis:** Data curation, Project administration, Funding acquisition. **Tanja Jovanovic:** Data curation, Project administration, Funding acquisition. **Judith K. Daniels:** Data curation, Project administration, Funding acquisition. **Martha Shenton:** Data curation, Project administration, Funding acquisition. **Nic J.A. van de Wee:** Data curation, Project administration, Funding acquisition. **Christian Schmahl:** Data curation, Project administration, Funding acquisition. **Milissa L. Kaufman:** Data curation, Project administration, Funding acquisition. **Isabelle M. Rosso:** Data curation, Project administration, Funding acquisition. **Scott R. Sponheim:** Data curation, Project administration, Funding acquisition. **David Bernd Hofmann:** Data curation, Project administration, Funding acquisition. **Richard A. Bryant:** Data curation, Project administration, Funding acquisition. **Kelene A. Fercho:** Data curation, Project administration, Funding acquisition. **Dan J. Stein:** Data curation, Project administration, Funding acquisition. **Sven C. Mueller:** Data curation, Project administration, Funding acquisition. **Bobak Hosseini:** Data curation, Project administration, Funding acquisition. **K. Luan Phan:** Data curation, Project administration, Funding acquisition. **Katie A. McLaughlin:** Data curation, Project administration, Funding acquisition. **Richard J. Davidson:** Data curation, Project administration, Funding acquisition. **Christine L. Larson:** Data curation, Project administration, Funding acquisition. **Geoffrey May:** Data curation, Project administration, Funding acquisition. **Steven M. Nelson:** Data curation, Project administration, Funding acquisition. **Chadi G. Abdallah:** Data curation, Project administration, Funding acquisition. **Hassaan Gomaa:** Data curation, Project administration, Funding acquisition. **Soraya Seedat:** Data curation, Project administration, Funding acquisition. **Ilan**

Harpaz-Rotem: Data curation, Project administration, Funding acquisition. **Israel Liberzon:** Data curation, Project administration, Funding acquisition. **Theo G.M. van Erp:** Data curation, Project administration, Funding acquisition. **Yann Quidé:** Data curation, Project administration, Funding acquisition. **Xin Wang:** Data curation, Project administration, Funding acquisition. **Paul M. Thompson:** Conceptualization, Methodology, Writing – review & editing. **Rajendra A. Morey:** Conceptualization, Methodology, Writing – review & editing, Investigation, Resources, Data curation, Supervision, Project administration, Funding acquisition.

Acknowledgments

DoD W81XWH-10-1-0925; Center for Brain and Behavior Research Pilot Grant; South Dakota Governor's Research Center Grant; CX001600 VA CDA; NHMRC Program Grant #1073041; R01 MH111671; VISN6 MIRECC; German Research Foundation grant to J. K. Daniels (DA 1222/4-1 and WA 1539/8-2); VA RR&D 1IK2RX000709; NIMH R01-MH043454; NIMH K01-MH122774; NIMH K01 MH118428-01 (Suarez-Jimenez); 5U01AA021681-08; K24MH71434; K24 DA028773; R01 MH63407; K99NS096116; VA RR&D 1K1RX002325; VA RR&D 1K2RX002922; MH101380; ZonMw, the Netherlands organization for Health Research and Development grant to Miranda Olff (40-00812-98-10041); Academic Medical Center Research Council grant to Miranda Olff (110614); VA CSR&D 1IK2CX001680; VISN17 Center of Excellence pilot funding; NIMH R01MH105535; NIMH 1R21MH102634; German Federal Ministry of Education and Research (BMBF RELEASE 01KR1303A); German Research Society (Deutsche Forschungsgemeinschaft, DFG; SFB/TRR 58: C06, C07); R01MH117601; R01AG059874; MJFF 14848; MH098212; MH071537; M01RR00039; UL1TR000454; HD071982; HD085850; R21MH112956; Anonymous Women's Health Fund; Kasparian Fund; Trauma Scholars Fund; Barlow Family Fund; NIMH K01 MH118467; W81XWH-08-2-0159; Department of Veterans Affairs via support for the National Center for PTSD; NIAAA via its support for (P50) Center for the Translational Neuroscience of Alcohol; NCATS via its support of (CTSA) Yale Center for Clinical Investigation; NIH R01 MH106574; F32MH109274; NIMH 1R21MH102634; R01MH113574; R01-MH103291; BOF 2–4 year project to Sven C. Mueller (01J05415); R01MH105355; Dana Foundation (to Dr. Nitschke); the University of Wisconsin Institute for Clinical and Translational Research; a National Science Foundation Graduate Research Fellowship (to Dr. Grupe); the National Institute of Mental Health (NIMH) R01 MH63407 (to De Bellis), R01 AA12479 (to De Bellis), and R01 MH61744 (to De Bellis); R01-MH043454 and T32-MH018931 (to Dr. Davidson); core grant to the Waisman Center from the National Institute of Child Health and Human Development (P30-HD003352); NIMH K23MH112873; Veterans Affairs Merit Review Program (10/01/08 – 09/30/13); L30 MH114379; South African Medical Research Council "SHARED ROOTS" Flagship Project; Grant -RFA-FSP-01-2013/SHARED ROOTS; South African Research Chair in PTSD from the Department of Science and Technology and the National Research Foundation; US Department of Defense Grant W81XWH08-2-0159 (PI: Stein, Murray B); VA RR&D I01RX000622; CDMRP W81XWH-08-2-0038; South African Medical Research Council; NARSAD Young Investigator; Department of Defense award number W81XWH-12-2-0012; ENIGMA was also supported in part by U54 EB020403 from the Big Data to Knowledge (BD2K) program; R56AG058854; R01MH116147;; P41 EB015922; 1R01MH110483; 1R21 MH098198; R01MH105355-01A. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs, the United States Government, or any other funding sources listed here.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119509.

References

Agarwal, P., Shroff, G., Malhotra, P., 2013. Approximate incremental big-data harmonization. In: 2013 IEEE International Congress on Big Data, pp. 118–125.

Beer, J.C., Tustison, N.J., Cook, P.A., Davatzikos, C., Sheline, Y.I., Shinohara, R.T., Linn, K.A., Alzheimer's Disease Neuroimaging Initiative, 2020. Longitudinal ComBat: a method for harmonizing longitudinal multi-scanner imaging data. Neuroimage 220, 117129.

Boedhoe, P.S., Schmaal, L., Abe, Y., Ameis, S.H., Arnold, P.D., Batistuzzo, M.C., Benedetti, F., Beucke, J.C., Bollettini, I., Bose, A., Brem, S., Calvo, A., Cheng, Y., Cho, K.I., Dallaspezia, S., Denys, D., Fitzgerald, K.D., Fouche, J.P., Gimenez, M., Gruner, P., Hanna, G.L., Hibar, D.P., Hoexter, M.Q., Hu, H., Huyser, C., Ikari, K., Jahanshad, N., Kathmann, N., Kaufmann, C., Koch, K., Kwon, J.S., Lazaro, L., Liu, Y., Lochner, C., Marsh, R., Martinez-Zalacain, I., Mataix-Cols, D., Menchon, J.M., Minuzzi, L., Nakamae, T., Nakao, T., Narayanaswamy, J.C., Piras, F., Piras, F., Pittenger, C., Reddy, Y.C., Sato, J.R., Simpson, H.B., Soreni, N., Soriano-Mas, C., Spalletta, G., Stevens, M.C., Szeszko, P.R., Tolin, D.F., Venkatasubramanian, G., Walitza, S., Wang, Z., van Wingen, G.A., Xu, J., Xu, X., Yun, J.Y., Zhao, Q., ENIGMA OCD Working Group, Thompson, P.M., Stein, D.J., van den Heuvel, O.A., 2017. Distinct subcortical volume alterations in pediatric and adult OCD: a worldwide meta- and mega-analysis. Am. J. Psychiatry 174, 60–69.

Brown, S.A., Brumback, T.Y., Tomlinson, K., Cummins, K., Thompson, W.K., Nagel, B.J., De Bellis, M.D., Hooper, S.R., Clark, D.B., Chung, T., Hasler, B.P., Colrain, I.M., Baker, F.C., Prouty, D., Pfefferbaum, A., Sullivan, E.V., Pohl, K.M., Rohlfing, T., Nichols, B.N., Chu, W.W., Tapert, S.F., 2015. The National Consortium on Alcohol and Neuro-Development in Adolescence (NCANDA): a multisite study of adolescent development and substance use. J. Stud. Alcohol Drugs 76, 895–908.

Davatzikos, C., 2019. Machine learning in neuroimaging: progress and challenges. Neuroimage 197, 652–656.

De Bellis, M.D., Hooper, S.R., Chen, S.D., Provenzale, J.M., Boyd, B.D., Glessner, C.E., MacFall, J.R., Payne, M.E., Rybczynski, R., Woolley, D.P., 2015. Posterior structural brain volumes differ in maltreated youth with and without chronic posttraumatic stress disorder. Dev. Psychopathol. 27, 1555–1576.

Dennis, E.L., Baron, D., Bartnik-Olson, B., Caeyenberghs, K., Esopenko, C., Hillary, F.G., Kenney, K., Koerte, I.K., Lin, A.P., Mayer, A.R., Mondello, S., Olsen, A., Thompson, P.M., Tate, D.F., Wilde, E.A., 2022. ENIGMA brain injury: framework, challenges, and opportunities. Hum. Brain Mapp. 43 (1), 149–166.

Destrieux, C., Fischl, B., Dale, A., Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 53, 1–15.

Dewey, B.E., Zhao, C., Reinhold, J.C., Carass, A., Fitzgerald, K.C., Sotirchos, E.S., Saidha, S., Oh, J., Pham, D.L., Calabresi, P.A., van Zijl, P.C.M., Prince, J.L., 2019. Deep harmony: a deep learning approach to contrast harmonization across scanner changes. Magn. Reson. Imaging 64, 160–170.

Favre, P., Pauling, M., Stout, J., Hozer, F., Sarrazin, S., Abe, C., Alda, M., Alloza, C., Alonso-Lana, S., Andreassen, O.A., Baune, B.T., Benedetti, F., Busatto, G.F., Canales-Rodriguez, E.J., Caseras, X., Chaim-Avancini, T.M., Ching, C.R.K., Dannlowski, U., Deppe, M., Eyler, L.T., Fatjo-Vilas, M., Foley, S.F., Grotegerd, D., Hajek, T., Haukvik, U.K., Howells, F.M., Jahanshad, N., Kugel, H., Lagerberg, T.V., Lawrie, S.M., Linke, J.O., McIntosh, A., Melloni, E.M.T., Mitchell, P.B., Polosan, M., Pomarol–Clotet, E., Repple, J., Roberts, G., Roos, A., Rosa, P.G.P., Salvador, R., Sarro, S., Schofield, P.R., Serpa, M.H., Sim, K., Stein, D.J., Sussmann, J.E., Temmingh, H.S., Thompson, P.M., Verdolini, N., Vieta, E., Wessa, M., Whalley, H.C., Zanetti, M.V., Leboyer, M., Mangin, J.F., Henry, C., Duchesnay, E., Houenou, J., ENIGMA Bipolar Disorder Working Group, 2019. Widespread white matter microstructural abnormalities in bipolar disorder: evidence from mega- and meta-analyses across 3033 individuals. Neuropsychopharmacology 44, 2285–2293.

Fortin, J.P., Cullen, N., Sheline, Y.I., Taylor, W.D., Aselcioglu, I., Cook, P.A., Adams, P., Cooper, C., Fava, M., McGrath, P.J., McInnis, M., Phillips, M.L., Trivedi, M.H., Weissman, M.M., Shinohara, R.T., 2018. Harmonization of cortical thickness measurements across scanners and sites. Neuroimage 167, 104–120.

Fortin, J.P., Parker, D., Tunc, B., Watanabe, T., Elliott, M.A., Ruparel, K., Roalf, D.R., Satterthwaite, T.D., Gur, R.C., Gur, R.E., Schultz, R.T., Verma, R., Shinohara, R.T., 2017. Harmonization of multi-site diffusion tensor imaging data. Neuroimage 161, 149–170.

Foster, E.D., Deardorff, A., 2017. Open Science Framework (OSF). J. Med. Libr. Assoc. 105, 203.

Frangou, S., Modabbernia, A., Williams, S.C.R., Papachristou, E., Doucet, G.E., Agartz, I., Aghajani, M., Akudjedu, T.N., Albajes-Eizagirre, A., Alnaes, D., Alpert, K.I., Andersson, M., Andreasen, N.C., Andreassen, O.A., Asherson, P., Banaschewski, T., Bargallo, N., Baumeister, S., Baur-Streubel, R., Bertolino, A., Bonvino, A., Boomsma, D.I., Borgwardt, S., Bourque, J., Brandeis, D., Breier, A., Brodaty, H., Brouwer, R.M., Buitelaar, J.K., Busatto, G.F., Buckner, R.L., Calhoun, V., Canales-Rodriguez, E.J., Cannon, D.M., Caseras, X., Castellanos, F.X., Cervenka, S., Chaim-Avancini, T.M., Ching, C.R.K., Chubar, V., Clark, V.P., Conrod, P., Conzelmann, A., Crespo-Facorro, B., Crivello, F., Crone, E.A., Dale, A.M., Dannlowski, U., Davey, C., de Geus, E.J.C., de Haan, L., de Zubicaray, G.I., den Braber, A., Dickie, E.W., Di Giorgio, A., Doan, N.T., Dorum, E.S., Ehrlich, S., Erk, S., Espeseth, T., Fatouros-Bergman, H., Fisher, S.E., Fouche, J.P., Franke, B., Frodl, T., Fuentes-Claramonte, P., Glahn, D.C., Gotlib, I.H., Grabe, H.J., Grimm, O., Groenewold, N.A., Grotegerd, D., Gruber, O., Gruner, P., Gur, R.E., Gur, R.C., Hahn, T., Harrison, B.J., Hartman, C.A., Hatton, S.N., Heinz, A., Heslenfeld, D.J., Hibar, D.P., Hickie, I.B., Ho, B.C., Hoekstra, P.J., Hohmann, S., Holmes, A.J., Hoogman, M., Hosten, N., Howells, F.M., Hulshoff Pol, H.E., Huyser, C., Jahanshad, N., James, A., Jernigan, T.L., Jiang, J., Jonsson, E.G., Joska, J.A., Kahn, R., Kalnin, A., Kanai, R., Klein, M., Klyushnik, T.P., Koenders, L., Koops, S., Kramer, B., Kuntsi, J., Lagopoulos, J., Lazaro, L., Lebedeva, I., Lee, W.H., Lesch, K.P., Lochner, C., Machielsen, M.W.J., Maingault, S., Martin, N.G.,

Martinez-Zalacain, I., Mataix-Cols, D., Mazoyer, B., McDonald, C., McDonald, B.C., McIntosh, A.M., McMahon, K.L., McPhilemy, G., Meinert, S., Menchon, J.M., Medland, S.E., Meyer-Lindenberg, A., Naaijen, J., Najt, P., Nakao, T., Nordvik, J.E., Nyberg, L., Oosterlaan, J., de la Foz, V.O., Paloyelis, Y., Pauli, P., Pergola, G., Pomarol–Clotet, E., Portella, M.J., Potkin, S.G., Radua, J., Reif, A., Rinker, D.A., Roffman, J.L., Rosa, P.G.P., Sacchet, M.D., Sachdev, P.S., Salvador, R., Sanchez-Juan, P., Sarro, S., Satterthwaite, T.D., Saykin, A.J., Serpa, M.H., Schmaal, L., Schnell, K., Schumann, G., Sim, K., Smoller, J.W., Sommer, I., Soriano-Mas, C., Stein, D.J., Strike, L.T., Swagerman, S.C., Tamnes, C.K., Temmingh, H.S., Thomopoulos, S.I., Tomyshev, A.S., Tordesillas-Gutierrez, D., Trollor, J.N., Turner, J.A., Uhlmann, A., van den Heuvel, O.A., van den Meer, D., van der Wee, N.J.A., van Haren, N.E.M., van 't Ent, D., van Erp, T.G.M., Veer, I.M., Veltman, D.J., Voineskos, A., Volzke, H., Walter, H., Walton, E., Wang, L., Wang, Y., Wassink, T.H., Weber, B., Wen, W., West, J.D., Westlye, L.T., Whalley, H., Wierenga, L.M., Wittfeld, K., Wolf, D.H., Worker, A., Wright, M.J., Yang, K., Yoncheva, Y., Zanetti, M.V., Ziegler, G.C., Karolinska Schizophrenia, P., Thompson, P.M., Dima, D., 2022. Cortical thickness across the lifespan: data from 17,075 healthy individuals aged 3-90 years. Hum. Brain Mapp. 43 (1), 431–451.

Hatton, S.N., Huynh, K.H., Bonilha, L., Abela, E., Alhusaini, S., Altmann, A., Alvim, M.K.M., Balachandra, A.R., Bartolini, E., Bender, B., Bernasconi, N., Bernasconi, A., Bernhardt, B., Bargallo, N., Caldairou, B., Caligiuri, M.E., Carr, S.J.A., Cavalleri, G.L., Cendes, F., Concha, L., Davoodi-bojd, E., Desmond, P.M., Devinsky, O., Doherty, C.P., Domin, M., Duncan, J.S., Focke, N.K., Foley, S.F., Gambardella, A., Gleichgerrcht, E., Guerrini, R., Hamandi, K., Ishikawa, A., Keller, S.S., Kochunov, P.V., Kotikalapudi, R., Kreilkamp, B.A.K., Kwan, P., Labate, A., Langner, S., Lenge, M., Liu, M., Lui, E., Martin, P., Mascalchi, M., Moreira, J.C.V., Morita-Sherman, M.E., O'Brien, T.J., Pardoe, H.R., Pariente, J.C., Ribeiro, L.F., Richardson, M.P., Rocha, C.S., Rodriguez-Cruces, R., Rosenow, F., Severino, M., Sinclair, B., Soltanian-Zadeh, H., Striano, P., Taylor, P.N., Thomas, R.H., Tortora, D., Velakoulis, D., Vezzani, A., Vivash, L., von Podewils, F., Vos, S.B., Weber, B., Winston, G.P., Yasuda, C.L., Zhu, A.H., Thompson, P.M., Whelan, C.D., Jahanshad, N., Sisodiya, S.M., McDonald, C.R., 2020. White matter abnormalities across different epilepsy syndromes in adults: an ENIGMA-Epilepsy Study. Brain 143, 2454–2473.

Hicks, R., Giacino, J., Harrison-Felix, C., Manley, G., Valadka, A., Wilde, E.A., 2013. Progress in developing common data elements for traumatic brain injury research: version two—the end of the beginning. J. Neurotrauma 30, 1852–1861.

Hofer, E., Roshchupkin, G.V., Adams, H.H.H., Knol, M.J., Lin, H.H., Li, S., Zare, H., Ahmad, S., Armstrong, N.J., Satizabal, C.L., Bernard, M., Bis, J.C., Gillespie, N.A., Luciano, M., Mishra, A., Scholz, M., Teumer, A., Xia, R., Jian, X.Q., Mosley, T.H., Saba, Y., Pirpamer, L., Seiler, S., Becker, J.T., Carmichael, O., Rotter, J.I., Psaty, B.M., Lopez, O.L., Amin, N., van der Lee, S.J., Yang, Q., Himali, J.J., Maillard, P., Beiser, A.S., DeCarli, C., Karama, S., Lewis, L., Harris, M., Bastin, M.E., Deary, I.J., Veronica Witte, A., Beyer, F., Loeffler, M., Mather, K.A., Schofield, P.R., Thalamuthu, A., Kwok, J.B., Wright, M.J., Ames, D., Trollor, J., Jiang, J.Y., Brodaty, H., Wen, W., Vernooij, M.W., Hofman, A., Uitterlinden, A.G., Niessen, W.J., Wittfeld, K., Bulow, R., Volker, U., Pausova, Z., Bruce Pike, G., Maingault, S., Crivello, F., Tzourio, C., Amouyel, P., Mazoyer, B., Neale, M.C., Franz, C.E., Lyons, M.J., Panizzon, M.S., Andreassen, O.A., Dale, A.M., Logue, M., Grasby, K.L., Jahanshad, N., Painter, J.N., Colodro-Conde, L., Bralten, J., Hibar, D.P., Lind, P.A., Pizzagalli, F., Stein, J.L., Thompson, P.M., Medland, S.E., Sachdev, P.S., Kremen, W.S., Wardlaw, J.M., Villringer, A., van Duijn, C.M., Grabe, H.J., Longstreth, W.T., Fornage, M., Paus, T., Debette, S., Ikram, M.A., Schmidt, H., Schmidt, R., Seshadri, S., Consortium, E., 2020. Genetic correlations and genome-wide associations of cortical structure in general population samples of 22,824 adults. Nat. Commun. 11 (1), 4796.

Johnson, W.E., Li, C., Rabinovic, A., 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8, 118–127.

Koshiyama, D., Miura, K., Nemoto, K., Okada, N., Matsumoto, J., Fukunaga, M., Hashimoto, R., 2022. Neuroimaging studies within Cognitive Genetics Collaborative Research Organization aiming to replicate and extend works of ENIGMA. Hum. Brain Mapp. 43 (1), 182–193.

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., Koopowitz, S., Levy, I., Nawijn, L., O'Connor, L., Olff, M., Salat, D.H., Sheridan, M.A., Spielberg, J.M., van Zuiden, M., Winternitz, S.R., Wolff, J.D., Wolf, E.J., Wang, X., Wrocklage, K., Abdallah, C.G., Bryant, R.A., Geuze, E., Jovanovic, T., Kaufman, M.L., King, A.P., Krystal, J.H., Lagopoulos, J., Bennett, M., Lanius, R., Liberzon, I., McGlinchey, R.E., McLaughlin, K.A., Milberg, W.P., Miller, M.W., Ressler, K.J., Veltman, D.J., Stein, D.J., Thomaes, K., Thompson, P.M., Morey, R.A., 2018. 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.

Mutlu, A.K., Schneider, M., Debbane, M., Badoud, D., Eliez, S., Schaer, M., 2013. Sex differences in thickness, and folding developments throughout the cortex. Neuroimage 82, 200–207.

Noble, S., Scheinost, D., Finn, E.S., Shen, X.L., Papademetris, X., McEwen, S.C., Bearden, C.E., Addington, J., Goodyear, B., Cadenhead, K.S., Mirzakhanian, H., Cornblatt, B.A., Olvet, D.M., Mathalon, D.H., McGlashan, T.H., Perkins, D.O., Belger, A., Seidman, L.J., Thermenos, H., Tsuang, M.T., van Erp, T.G.M., Walker, E.F., Hamann, S., Woods, S.W., Cannon, T.D., Constable, R.T., 2017. Multisite reliability of MR-based functional connectivity. Neuroimage 146, 959–970.

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.J., Volzke, H., Johnson, S.C., Fripp, J., Koutsouleris, N., Wolf, D.H., Gur, R., Gur, R., Morris, J., Albert, M.S., Grabe, H.J., Resnick, S.M., Bryan, R.N., Wolk, D.A., Shinohara, R.T., Shou, H.C., Davatzikos, C., 2020. Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. Neuroimage 208, 116450.

Radua, J., Vieta, E., Shinohara, R., Kochunov, P., Quide, Y., Green, M.J., Weickert, C.S., Weickert, T., Bruggemann, J., Kircher, T., Nenadic, I., Cairns, M.J., Seal, M., Schall, U., Henskens, F., Fullerton, J.M., Mowry, B., Pantelis, C., Lenroot, R., Cropley, V., Loughland, C., Scott, R., Wolf, D., Satterthwaite, T.D., Tan, Y., Sim, K., Piras, F., Spalletta, G., Banaj, N., Pomarol-Clotet, E., Solanes, A., Albajes-Eizagirre, A., Canales-Rodriguez, E.J., Sarro, S., Di Giorgio, A., Bertolino, A., Stablein, M., Oertel, V., Knochel, C., Borgwardt, S., du Plessis, S., Yun, J.Y., Kwon, J.S., Dannlowski, U., Hahn, T., Grotegerd, D., Alloza, C., Arango, C., Janssen, J., Diaz-Caneja, C., Jiang, W., Calhoun, V., Ehrlich, S., Yang, K., Cascella, N.G., Takayanagi, Y., Sawa, A., Tomyshev, A., Lebedeva, I., Kaleda, V., Kirschner, M., Hoschl, C., Tomecek, D., Skoch, A., van Amelsvoort, T., Bakker, G., James, A., Preda, A., Weideman, A., Stein, D.J., Howells, F., Uhlmann, A., Temmingh, H., Lopez-Jaramillo, C., Diaz-Zuluaga, A., Fortea, L., Martinez-Heras, E., Solana, E., Llufriu, S., Jahanshad, N., Thompson, P., Turner, J., van Erp, T.ENIGMA Consortium collaborators, 2020. Increased power by harmonizing structural MRI site differences with the ComBat batch adjustment method in ENIGMA. Neuroimage 218, 116956.

Rast, P., Kennedy, K.M., Rodrigue, K.M., Robinson, P., Gross, A.L., McLaren, D.G., Grabowski, T., Schaie, K.W., Willis, S.L., 2018. APOEepsilon4 genotype and hypertension modify 8-year cortical thinning: five occasion evidence from the Seattle Longitudinal Study. Cereb. Cortex 28, 1934–1945.

Ritchie, S.J., Cox, S.R., Shen, X., Lombardo, M.V., Reus, L.M., Alloza, C., Harris, M.A., Alderson, H.L., Hunter, S., Neilson, E., Liewald, D.C.M., Auyeung, B., Whalley, H.C., Lawrie, S.M., Gale, C.R., Bastin, M.E., McIntosh, A.M., Deary, I.J., 2018. Sex differences in the adult human brain: evidence from 5216 UK Biobank participants. Cereb. Cortex 28, 2959–2975.

Santhanam, P., Wilson, S.H., Mulatya, C., Oakes, T.R., Weaver, L.K., 2019. Age-accelerated reduction in cortical surface area in United States service members and veterans with mild traumatic brain injury and post-traumatic stress disorder. J. Neurotrauma 36, 2922–2929.

Savalia, N.K., Agres, P.F., Chan, M.Y., Feczko, E.J., Kennedy, K.M., Wig, G.S., 2017. Motion-related artifacts in structural brain images revealed with independent estimates of in-scanner head motion. Hum. Brain Mapp. 38, 472–492.

Savjani, R.R., Taylor, B.A., Acion, L., Wilde, E.A., Jorge, R.E., 2017. Accelerated changes in cortical thickness measurements with age in military service members with traumatic brain injury. J. Neurotrauma 34, 3107–3116.

Shalev, A., Liberzon, I., Marmar, C., 2017. Post-traumatic stress disorder. N. Engl. J. Med. 376, 2459–2469.

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., Bulow, R., Chen, J., Chye, Y., Dannlowski, U., de Kovel, C.G.F., Donohoe, G., Eyler, L.T., Faraone, S.V., Favre, P., Filippi, C.A., Frodl, T., Garijo, D., Gil, Y., Grabe, H.J., Grasby, K.L., Hajek, T., Han, L.K.M., Hatton, S.N., Hilbert, K., Ho, T.F.C., Holleran, L., Homuth, G., Hosten, N., Houenou, J., Ivanov, I., Jia, T.Y., Kelly, S., Klein, M., Kwon, J.S., Laansma, M.A., Leerssen, J., Lueken, U., Nunes, A., Neill, J.O., Opel, N., Piras, F., Piras, F., Postema, M.C., Pozzi, E., Shatokhina, N., Soriano-Mas, C., Spalletta, G., Sun, D.Q., Teumer, A., Tilot, A.K., Tozzi, L., van der Merwe, C., Van Someren, E.J.W., van Wingen, G.A., Volzke, H., Walton, E., Wang, L., Winkler, A.M., Wittfeld, K., Wright, M.J., Yun, J.Y., Zhang, G.H., Zhang-James, Y., Adhikari, B.M., Agartz, I., Aghajani, M., Aleman, A., Althoff, R.R., Altmann, A., Andreassen, O.A., Baron, D.A., Bartnik-Olson, B.L., Bas-Hoogendam, J.M., Baskin-Sommers, A.R., Bearden, C.E., Berner, L.A., Boedhoe, P.S.W., Brouwer, R.M., Buitelaar, J.K., Caeyenberghs, K., Cecil, C.A.M., Cohen, R.A., Cole, J.H., Conrod, P.J., De Brito, S.A., de Zwarte, S.M.C., Dennis, E.L., Desrivieres, S., Dima, D., Ehrlich, S., Esopenko, C., Fairchild, G., Fisher, S.E., Fouche, J.P., Francks, C., Frangou, S., Franke, B., Garavan, H.P., Glahn, D.C., Groenewold, N.A., Gurholt, T.P., Gutman, B.A., Hahn, T., Harding, I.H., Hernaus, D., Hibar, D.P., Hillary, F.G., Hoogman, M., Pol, H.H.E., Jalbrzikowski, M., Karkashadze, G.A., Klapwijk, E.T., Knickmeyer, R.C., Kochunov, P., Koerte, I.K., Kong, X.Z., Liew, S.L., Lin, A.L.P., Logue, M.W., Luders, E., Macciardi, F., Mackey, S., Mayer, A.R., McDonald, C.R., McMahon, A.B., Medland, S.E., Modinos, G., Morey, R.A., Mueller, S.C., Mukherjee, P., Namazova-Baranova, L., Nir, T.M., Olsen, A., Paschou, P., Pine, D.S., Pizzagalli, F., Renteria, M.E., Rohrer, J.D., Samann, P.G., Schmaal, L., Schumann, G., Shiroishi, M.S., Sisodiya, S.M., Smit, D.J.A., Sonderby, I.E., Stein, D.J., Stein, J.L., Tahmasian, M., Tate, D.F., Turner, J.A., van den Heuvel, O.A., van der Wee, N.J.A., van der Werf, Y.D., van Erp, T.G.M., van Haren, N.E.M., van Rooij, D., van Velzen, L.S., Veer, I.M., Veltman, D.J., Villalon-Reina, J.E., Walter, H., Whelan, C.D., Wilde, E.A., Zarei, M., Zelman, V.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. Transl. Psychiatry 10 (1), 100.

van Rooij, D., Anagnostou, E., Arango, C., Auzias, G., Behrmann, M., Busatto, G.F., Calderoni, S., Daly, E., Deruelle, C., Di Martino, A., Dinstein, I., Duran, F.L.S., Durston, S., Ecker, C., Fair, D., Fedor, J., Fitzgerald, J., Freitag, C.M., Gallagher, L., Gori, I., Haar, S., Hoekstra, L., Jahanshad, N., Jalbrzikowski, M., Janssen, J., Lerch, J., Luna, B., Martinho, M.M., McGrath, J., Muratori, F., Murphy, C.M., Murphy, D.G.M., O'Hearn, K., Oranje, B., Parellada, M., Retico, A., Rosa, P., Rubia, K., Shook, D., Taylor, M., Thompson, P.M., Tosetti, M., Wallace, G.L., Zhou, F., Buitelaar, J.K., 2018. Cortical and subcortical brain morphometry differences between patients with autism spectrum disorder and healthy individuals across the lifespan: results from the ENIGMA ASD Working Group. Am. J. Psychiatry 175, 359–369.

Volkow, N.D., Koob, G.F., Croyle, R.T., Bianchi, D.W., Gordon, J.A., Koroshetz, W.J., Perez-Stable, E.J., Riley, W.T., Bloch, M.H., Conway, K., Deeds, B.G., Dowling, G.J., Grant, S., Howlett, K.D., Matochik, J.A., Morgan, G.D., Murray, M.M., Noronha, A., Spong, C.Y., Wargo, E.M., Warren, K.R., Weiss, S.R.B., 2018. The conception of the

ABCD study: from substance use to a broad NIH collaboration. Dev. Cogn. Neurosci. 32, 4–7.

Walhovd, K.B., Fjell, A.M., Giedd, J., Dale, A.M., Brown, T.T., 2017. Through thick and thin: a need to reconcile contradictory results on trajectories in human cortical development. Cereb. Cortex 27, 1472–1481.

Wang, X., Xie, H., Chen, T., Cotton, A.S., Salminen, L.E., Logue, M.W., Clarke-Rubright, E.K., Wall, J., Dennis, E.L., O'Leary, B.M., Abdallah, C.G., Andrew, E., Baugh, L.A., Bomyea, J., Bruce, S.E., Bryant, R., Choi, K., Daniels, J.K., Davenport, N.D., Davidson, R.J., DeBellis, M., deRoon-Cassini, T., Disner, S.G., Fani, N., Fercho, K.A., Fitzgerald, J., Forster, G.L., Frijling, J.L., Geuze, E., Gomaa, H., Gordon, E.M., Grupe, D., Harpaz-Rotem, I., Haswell, C.C., Herzog, J.I., Hofmann, D., Hollifield, M., Hosseini, B., Hudson, A.R., Ipser, J., Jahanshad, N., Jovanovic, T., Kaufman, M.L., King, A.P., Koch, S.B.J., Koerte, I.K., Korgaonkar, M.S., Krystal, J.H., Larson, C., Lebois, L.A.M., Levy, I., Li, G., Magnotta, V.A., Manthey, A., May, G., McLaughlin, K.A., Mueller, S.C., Nawijn, L., Nelson, S.M., Neria, Y., Nitschke, J.B., Olff, M., Olson, E.A., Peverill, M., Luan Phan, K., Rashid, F.M., Ressler, K., Rosso, I.M., Sambrook, K., Schmahl, C., Shenton, M.E., Sierk, A., Simons, J.S., Simons, R.M., Sponheim, S.R., Stein, M.B., Stein, D.J., Stevens, J.S., Straube, T., Suarez-Jimenez, B., Tamburrino, M., Thomopoulos, S.I., van der Wee, N.J.A., van der Werff, S.J.A., van Erp, T.G.M., van Rooij, S.J.H., van Zuiden, M., Varkevisser, T., Veltman, D.J., Vermeiren, R., Walter, H., Wang, L., Zhu, Y., Zhu, X., Thompson, P.M., Morey, R.A., Liberzon, I., 2021. Cortical volume abnormalities in posttraumatic stress disorder: an ENIGMA-psychiatric genomics consortium PTSD workgroup mega-analysis. Mol. Psychiatry 26 (8), 4331–4343.

Yu, M.C., Linn, K.A., Cook, P.A., Phillips, M.L., McInnis, M., Fava, M., Trivedi, M.H., Weissman, M.M., Shinohara, R.T., Sheline, Y.I., 2018. Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. Hum. Brain Mapp. 39, 4213–4227.

Zugman, A., Harrewijn, A., Cardinale, E.M., Zwiebel, H., Freitag, G.F., Werwath, K.E., Bas-Hoogendam, J.M., Groenewold, N.A., Aghajani, M., Hilbert, K., Cardoner, N., Porta-Casteras, D., Gosnell, S., Salas, R., Blair, K.S., Blair, J.R., Hammoud, M.Z., Milad, M., Burkhouse, K., Phan, K.L., Schroeder, H.K., Strawn, J.R., Beesdo-Baum, K., Thomopoulos, S.I., Grabe, H.J., Van der Auwera, S., Wittfeld, K., Nielsen, J.A., Buckner, R., Smoller, J.W., Mwangi, B., Soares, J.C., Wu, M.J., Zunta-Soares, G.B., Jackowski, A.P., Pan, P.M., Salum, G.A., Assaf, M., Diefenbach, G.J., Brambilla, P., Maggioni, E., Hofmann, D., Straube, T., Andreescu, C., Berta, R., Tamburo, E., Price, R., Manfro, G.G., Critchley, H.D., Makovac, E., Mancini, M., Meeten, F., Ottaviani, C., Agosta, F., Canu, E., Cividini, C., Filippi, M., Kostic, M., Munjiza, A., Filippi, C.A., Leibenluft, E., Alberton, B.A.V., Balderston, N.L., Ernst, M., Grillon, C., Mujica-Parodi, L.R., van Nieuwenhuizen, H., Fonzo, G.A., Paulus, M.P., Stein, M.B., Gur, R.E., Gur, R.C., Kaczkurkin, A.N., Larsen, B., Satterthwaite, T.D., Harper, J., Myers, M., Perino, M.T., Yu, Q., Sylvester, C.M., Veltman, D.J., Lueken, U., Van der Wee, N.J.A., Stein, D.J., Jahanshad, N., Thompson, P.M., Pine, D.S., Winkler, A.M., 2022. Mega–analysis methods in ENIGMA: the experience of the generalized anxiety disorder working group. Hum. Brain Mapp. 43 (1), 255–277.