ORIGINAL ARTICLE

Genomics of posttraumatic stress disorder in veterans: Methods and rationale for Veterans Affairs Cooperative Study #575B

Krishnan Radhakrishnan^{1,2} | Mihaela Aslan^{1,3} | Kelly M. Harrington^{4,5} | Robert H. Pietrzak^{1,6} | Grant Huang⁷ | Sumitra Muralidhar⁷ | Kelly Cho⁴ | Rachel Quaden⁴ | David Gagnon^{4,8} | Saiju Pyarajan⁴ | Ning Sun^{1,3} | Hongyu Zhao^{1,3} | Michael Gaziano^{4,9} | John Concato^{1,3} | Murray B. Stein^{10,11} | Joel Gelernter^{3,12}

¹Clinical Epidemiology Research Center (CERC), VA Connecticut Healthcare System, West Haven, Connecticut, USA

²College of Medicine, University of Kentucky, Lexington, Kentucky, USA

³School of Medicine, Yale University, New Haven, Connecticut, USA

⁴Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, Massachusetts, USA

⁵School of Medicine, Boston University, Boston, Massachusetts, USA

⁶U.S. Department of Veterans Affairs National Center for Posttraumatic Stress Disorder, Clinical Neurosciences Division, VA Connecticut Healthcare System, West Haven, Connecticut, USA

⁷Office of Research and Development, Veterans Health Administration, Washington, DC, USA

⁸School of Public Health, Boston University, Boston, Massachusetts, USA

⁹Harvard Medical School, Harvard University, Boston, Massachusetts, USA

¹⁰VA San Diego Healthcare System, San Diego, California, USA

¹¹School of Medicine, University of California, San Diego, La Jolla, California, USA

¹²Psychiatry Service, VA Connecticut Healthcare System, West Haven, Connecticut, USA

Correspondence

Krishnan Radhakrishnan, Clinical Epidemiology Research Center, VA Connecticut Healthcare System, 151B, 950 Campbell Ave., West Haven, CT 06516, USA.
Email: krishnan.radhakrishnan@va.gov

Funding information

Department of Veterans Affairs Cooperative Studies Program

Abstract

Objectives: Heritability in the risk for developing posttraumatic stress disorder (PTSD) has been established, but most genome-wide association studies (GWASs) of PTSD involve relatively small sample sizes and limited identification of associated genetic loci. This report describes the methodology of a Veterans Affairs (VA) Cooperative Studies Program GWAS of PTSD among combat-exposed U.S. veterans.

Methods: Probable cases (with PTSD) and probable controls (without PTSD) were identified from among veterans enrolled in the VA Million Veteran Program (MVP) with an algorithm developed using questionnaire responses and electronic health record information. This algorithm, based on a statistical model, relied on medical chart reviews as a reference standard and was refined using telephone interviews. Subsequently, to evaluate the impact of probabilistic phenotyping on statistical power, the threshold probability for case-control selection was varied in simulations.

Results: As of September 2018, >695,000 veterans have enrolled in MVP. For current analyses, genotyping data were available for >353,000 participants, including >83,000 combat-exposed veterans. A threshold probability of 0.7 for case and control designation yielded an interim >16,000 cases and >33,000 controls.

Conclusions: A formal methodological approach was used to identify cases and controls for subsequent GWAS analyses to identify genetic risk loci for PTSD.

KEYWORDS

Clinician-Administered PTSD Scale, combat-exposed veterans, genomics, medical record review, posttraumatic stress disorder (PTSD)

This article has been contributed to by US Government employees and their work is in the public domain in the USA.

1 | INTRODUCTION

Posttraumatic stress disorder (PTSD), classified by the *Diagnostic and Statistical Manual of Mental Disorders 5* (DSM-5) among trauma and stress-related disorders, can be a severe, debilitating condition marked by distress and impairment in life tasks (Marmar et al., 2015; Sareen, 2014). Reports have suggested associations of PTSD with increased risk for psychiatric comorbidity, substance use disorders, chronic pain, suicidality, cardiovascular disease, obesity, metabolic syndrome, and dementia (Gros, Szafranski, Brady, & Back, 2015; Mawanda, Wallace, McCoy, & Abrams, 2017; Sareen, 2014; Stein et al., 2017). The disorder can develop after exposure to a stressful or traumatic event, such as serious injury, sexual assault, or threatened death (Gates et al., 2012). Conversely, many people are resilient to such exposures (Jang, Taylor, Stein, & Yamagata, 2007).

DSM-5 PTSD diagnostic criteria require exposure to a stressor or traumatic event, and perseverance for ≥ 1 month of symptoms from four core categories: vivid re-experiencing intrusions, active and/or passive avoidance, negative cognitions and mood, and hyperarousal (Sareen, 2014). Stress-related symptoms among those who develop PTSD may or may not resolve; symptoms can last for long periods of time and may increase in severity (Xie et al., 2013).

Military combat exposure and sexual trauma are strong risk factors for developing PTSD (Jorge, 2015; Phillips, Leardmann, Gumbs, & Smith, 2010; Zamorski & Boulos, 2014). Thus, although the prevalence of PTSD in the U.S. civilian population is approximately 6% (Goldstein et al., 2016), the lifetime prevalence of PTSD in combat-exposed U.S. veterans is estimated to be as high as 32% (Wisco et al., 2014). Vietnam combat veterans reportedly have a lifetime PTSD prevalence of 19–31% (Gates et al., 2012; Murphy, Iversen, & Greenberg, 2008), with persistent/chronic PTSD estimated at 5–17% (Gates et al., 2012; Marmar et al., 2015; Murphy et al., 2008). The PTSD prevalence among combat-exposed veterans deployed to Iraq or Afghanistan varies widely, depending on the military branch and PTSD definition used—a meta-analysis of 33 studies, including nearly 5 million veterans, reported a prevalence of 23% (Fulton et al., 2015).

PTSD is influenced by genetic factors, as demonstrated by twin and other genetic epidemiology studies (Pitman et al., 2006; Rodgers & Bale, 2015; Smoller, 2016). Knowledge of the specifics of these genetic factors would yield a variety of benefits, ranging from improved biological understanding and better ability to differentiate disorders that are genetically related from those that are genetically distinct, to discovering possible pharmacological targets (Stein, 2018; Stein & Smoller, 2018). The best method for identifying these genetic risk factors is through genome-wide approaches, such as genome-wide association studies (GWASs). In a little over a decade, GWASs, when adequately powered, have yielded genetic clues across essentially the entire spectrum of genetically complex traits including mental health disorders (Sullivan et al., 2018).

Several GWASs of PTSD have been conducted recently (Almli et al., 2015; Ashley-Koch et al., 2015; Daskalakis, Rijal, King, Huckins, & Ressler, 2018; Duncan et al., 2018; Guffanti et al., 2013; Logue et al., 2013; Nievergelt et al., 2015; Nievergelt et al., 2018; Stein et al., 2016; Xie et al., 2013), including in veteran populations (Ashley-Koch

et al., 2015). The sample sizes in GWASs to date have varied, with the largest being a meta-analysis of 20,730 persons (Duncan et al., 2018). Replication of primary risk loci for PTSD has been a challenge, presumably because of small sizes of discovery and replication samples (Banerjee, Morrison, & Ressler, 2017; Sheerin, Lind, Bountress, Nugent, & Amstadter, 2017).

The current project investigates the genetics of PTSD using the Veterans Affairs (VA)-based Million Veteran Program (MVP; Gaziano et al., 2016), designed to enable study of how genes affect health by building a mega-biobank of DNA samples and health information. The VA also has a long-established history of successfully using electronic health record (EHR) systems in research (Brown, Lincoln, Groen, & Kolodner, 2003), providing an unparalleled opportunity to link participants' genetic data to clinical outcomes.

The current study, "Genomics of Posttraumatic Stress Disorder in Veterans," supported by the VA Cooperative Studies Program (CSP) as CSP study #575B, is focused on a GWAS of PTSD in a large sample of combat-exposed U.S. veterans. Recognizing considerable uncertainty associated with PTSD diagnosis when determined from EHRs alone (Holowka et al., 2014; Ouyang, Apley, & Mehrotra, 2016), a probabilistic methodology (Sinnott et al., 2014) is used to estimate PTSD caseness and noncaseness (Harrington et al., in press) and to account for phenotypic misclassification in the power and sample size calculations for the GWAS.

2 | METHODS

2.1 | Overview

This study was approved by the Institutional Review Boards at VA Medical Centers in Boston, MA, San Diego, CA, and West Haven, CT. Using a case-control design, genotype is the main explanatory variable, and several phenotype definitions (trait formulations), including lifetime PTSD diagnosis, are the outcome variables. The primary analysis will focus on combat-exposed PTSD cases and combat-exposed non-PTSD controls, selected from >353,000 genotyped participants in the MVP. The restriction to combat-exposed cases and combat-exposed controls promotes an unbiased comparison, even if certain participants have noncombat (e.g., early-life and later-in-life) trauma that may have caused or contributed to their PTSD.

2.2 | Algorithm for selecting cases and controls

Inclusion criteria for selection were history of military deployment and combat exposure. The only exclusion criterion was evidence of schizophrenia or bipolar disorder. Information was obtained from a combination of self-reports in the MVP Baseline and Lifestyle questionnaires (Gaziano et al., 2016), together with clinical and administrative data from the VA EHR.

From a pragmatic perspective, given the projected sample size in the tens of thousands, automated PTSD phenotyping by means of an EHR-based algorithm would be more practicable than face-to-face or telephone interviews, which are both resource intensive and time-consuming. For example, telephone interviews require approximately

1 hr per participant (not including time for preinterview arrangements and postinterview analyses), with the added burden of being potentially a source of psychological discomfort to study participants. Thus, given the size of the MVP participant pool, feasibility and affordability become major, if not prohibitive, concerns. Conversely, EHR-based diagnoses do not involve patient contact, and after algorithms have been properly calibrated and are shown to be sufficiently robust, large populations can be analyzed in little time and at low cost.

Potential cases and controls were therefore identified from the VA EHR using the algorithm summarized below (see Figure 1) and described elsewhere (Harrington et al., in press). To develop and assess its predictive capability, reference data were generated by identifying confirmed PTSD cases and controls from 500 in-depth VA EHR reviews, conducted by five experts in PTSD diagnosis at VA Medical Centers in Boston, MA, San Diego, CA, and West Haven, CT. Evidence of PTSD from the DSM-IV PTSD Checklist (PCL) score (Forbes, Creamer, & Biddle, 2001) on the MVP Lifestyle Survey was also used, when available. Veterans who met some PTSD criteria, but could not clearly be classified as cases or controls, were considered “possible” cases. Finally, 125 (25%) of the charts were independently rated by two reviewers to establish interrater reliability.

Based on these chart review results as the reference standard, and using the same three PTSD classes, a probabilistic-based algorithm was constructed for automated selection of PTSD cases and controls from the MVP source population (Harrington et al., in press). In particular, symmetric multinomial logistic regression was used, which models a participant's conditional probability π_k of “belonging” to class k (=case, control, and possible case) given the predictor $\mathbf{x}$ for the participant, that is, $\Pr(\text{PTSD status} = k | \mathbf{x})$, expressed as (Huttunen, Manninen, Kauppi, & Tohka, 2013)

$$\pi_k(\mathbf{x}) = \frac{\exp(\beta_{k0} + \beta_k^T \mathbf{x})}{\sum_{j=1}^K \exp(\beta_{j0} + \beta_j^T \mathbf{x})}, \quad k = 1, \dots, K, \quad (1)$$

where the p -dimensional vector $\mathbf{x}$ contains the covariates of interest, K is the number of classes, and β_{k0} and $\beta_k = (\beta_{k1}, \beta_{k2}, \dots, \beta_{kp})^T$ are the coefficients of the model for the k th class. The covariates were selected based on their potential association with PTSD. The regression coefficients were estimated using the least absolute shrinkage and selection operator (LASSO) penalized multinomial logistic regression, with 10-fold cross validation. This method simultaneously selects predictors to be retained and estimates the corresponding regression coefficients $\{\beta_{kj}\}$ (Wu, Chen, Hastie, Sobel, & Lange, 2009).

After the best fit model is derived, Equation (1) gives the participant's estimated probability of belonging to each of the three PTSD classes (Figure 1). Considerable flexibility exists regarding use of these probabilities to assign PTSD status. For example, the predicted class $\hat{k}$ could be defined as the maximum of the three probabilities, that is, $\hat{k} = \arg\max_k \pi(k|\mathbf{x})$. The results obtained using this definition are reported in Harrington et al. (in press).

FIGURE 1 Schematic of algorithm for selecting PTSD cases and controls from the VA EHR. EHR: electronic health record; ICD: *International Classification of Diseases*; PCL: PTSD Checklist; PTSD: posttraumatic stress disorder; VHA: Veterans Health Administration

For application to CSP#575B, a more conservative rule was adopted to minimize false-positive and false-negative designations. Specifically, $\hat{k}$ is defined as the class for which the probability is greater than a minimal or threshold value, $\pi_{\min}$, which necessarily must be ≥ 0.5 to ensure uniqueness of the prediction. Thus, $\hat{k} = k$ such that $\pi_k > \pi_{\min} \geq 0.5$, where $\pi_{\min}$ was optimized as discussed below.

2.3 | Optimizing $\pi_{\min}$ for case-control GWAS

The optimal threshold probability (π_{opt}), defined as the $\pi_{\min}$ that maximized the statistical power to detect association between single nucleotide polymorphisms (SNPs) and disease (PTSD), was obtained as follows. For a given $\pi_{\min}$, the power was computed from the predicted number of PTSD cases and controls obtained with Equation (1) and the uncertainties in phenotyping, assuming multiplicative genetic effects and PTSD prevalence of 0.3 (Hou et al., 2017). The minor allele frequency (MAF) was varied from 0.10 to 0.50 and relative risk (RR) for disease-associated SNPs from 1.05 to 1.20. Starting with a value

of 0.60, $\pi_{\min}$ was increased in increments of 0.05, the above procedure repeated, and the $\pi_{\min}$ that maximized power selected. To further minimize false-negative designations and thus increase reliability of predicted controls, we excluded participants predicted to be controls by the algorithm if their PCL scores were greater than 30, inasmuch as such scores suggest likely PTSD case status. For example, PCL scores of 33–39 were previously found to be optimally efficient for identifying different classifications of PTSD, from partial to full PTSD (Dickstein et al., 2015).

2.4 | Sensitivity analyses

In planned GWAS analyses, although our primary analysis will classify PTSD as a binary yes/no trait, we will also consider PTSD as an ordinal trait, based on the probability of PTSD caseness; this reclassification should improve statistical power. Other analyses will help determine the degree to which comorbidity or sub-components of the PTSD syndrome (e.g., hyperarousal, re-experiencing, sleep disorder, and alcohol use disorder) appears to explain particular genetic findings.

2.5 | Telephone interviews (CAPS-5)

To validate and potentially refine the algorithm, clinicians are conducting telephone interviews in a sample of algorithm-predicted probable cases and probable controls, using the DSM-5-compatible version of the Clinician-Administered PTSD Scale (CAPS-5; Weathers et al., 2018). Only participants predicted by the algorithm to be cases ($\text{Pr}(\text{case}) > \pi_{\text{opt}}$) or controls ($\text{Pr}(\text{control}) > \pi_{\text{opt}}$) are selected to be interviewed; those predicted to be possible cases because $\text{Pr}(\text{case}) \leq \pi_{\text{opt}}$ and $\text{Pr}(\text{control}) \leq \pi_{\text{opt}}$ are excluded. Telephone interviews using the CAPS instrument have been shown to perform well relative to face-to-face interviews for PTSD (Aziz & Kenford, 2004; Prescott et al., 2014). The interviewers, based at the San Diego VA Medical Center, are blind to participants' diagnosis and algorithm assignment. Interviewers have extensive experience with use of the CAPS (including training by Frank Weathers, PhD, a developer of the CAPS) and with patients diagnosed with PTSD or other psychiatric disorders. To assess concordance of the diagnoses based on DSM-IV criteria and on DSM-5 criteria, a random subset of cases and controls per chart review (primarily DSM-IV criteria) are being selected for telephone interview (DSM-5 criteria). In addition, although the interview uses the CAPS-5, which adheres to DSM-5 criteria for PTSD, questions also pertain to DSM-IV criteria. The use of this interview enables us to determine if an individual meets criteria for PTSD using DSM-IV, DSM-5, neither, or both. Interrater reliability will be assessed by a second clinician's evaluation of a randomly selected subsample (approximately 10%) of interviews.

2.6 | Genotyping

Details regarding collection, shipping, and storage of biospecimens, and specifics of the MVP microarray, have been described previously (Gaziano et al., 2016). In brief, blood samples from consented participants are mailed to the VA Central Biorepository in Boston;

genotyping is done by two vendors, BioStorage Technologies, Inc. (Indianapolis, IN), and AKESOGen (Norcross, GA), using a customized Affymetrix® 723K chip. After basic quality control, data are maintained by the VA's Genomic Information System for Integrated Science. Imputation was performed on the called genotype data using the 1000 Genomes Project phase 3 reference panel (Genomes Project Consortium et al., 2015) using Minimac3 (Howie, Fuchsberger, Stephens, Marchini, & Abecasis, 2012).

2.7 | Data analysis

In planning the study, sample size calculations were performed using the CaTS software (Skol, Scott, Abecasis, & Boehnke, 2006). We used a one-stage design, an alpha level of 5×10^{-8} for each marker, and a multiplicative inheritance model. Both dominant and recessive disease inheritance models were considered, and MAF was varied between 0.1 and 0.5. Prior to having access to MVP data, we projected (conservatively) a PTSD prevalence of 9%, as well as sample sizes of 10,000 cases (as the limiting factor) and >10,000 controls. For these assumptions and statistical power of 80%, the minimal odds ratio for discovering association between SNPs and PTSD ranged between 1.12 and 1.97.

Association tests will be performed using logistic regression models implemented in the PLINK software package (<http://pngu.mgh.harvard.edu/~purcell/plink/>), adjusted for age, sex, and the principal components of ancestry, determined with the Eigensoft package (<http://www.hsph.harvard.edu/alkes-price/software/>). Other options include fastPCA (Galinsky et al., 2016) to infer principal components and RVTESTS (Zhan, Hu, Li, Abecasis, & Liu, 2016) for association analysis. Genetic effects will be modeled as additive effects of each copy of the minor allele. Because our samples will be genetically heterogeneous, European American and African American participants will be analyzed separately. European Americans and African Americans will also be meta-analyzed together to uncover additional support for risk loci shared between these populations and for fine mapping.

In addition to SNP-based analysis, we will incorporate prior biological knowledge (e.g., pathway information, gene expression networks, and interaction networks) to identify genes and SNPs associated with PTSD and related phenotypes. These gene-based results can be summarized to identify pathways enriched for genes showing association signals (Ballard, Abraham, Cho, & Zhao, 2010). Pathway topological information may also be informative regarding genetic associations (Chen, Cho, & Zhao, 2011). We will consider functional impact of putative causal variants as described elsewhere (Lu, Powles, Wang, He, & Zhao, 2016). We will also explore gene–gene ($G \times G$) and gene–environment ($G \times E$) interactions, including pathway-based interaction analysis to reduce the search space (Chen et al., 2011), and gene–environment interaction at a genome-wide level (Polimanti et al., 2018).

3 | RESULTS

3.1 | MVP participants' demographics and health characteristics

Between launch of the MVP in 2011 and September 2018, over 4.5 million invitations were mailed to eligible veterans, and walk-ins

TABLE 1 Demographic and military service-related characteristics of genotyped Million Veteran Program enrollees with completed Baseline and Lifestyle surveys ($N = 173,574$)

Factor	Frequency (%) or mean (SD)
Demographic factors	
Age (years)	
18–29	1,198 (0.7)
30–39	3,769 (2.2)
40–49	9,321 (5.4)
50–59	25,425 (14.6)
60–69	74,476 (42.9)
70–79	37,173 (21.4)
80+	22,211 (12.8)
Missing	1 (0.0)
Mean (SD)	65.9 (11.7)
Median	66
Sex	
Male	160,447 (92.4)
Female	13,126 (7.6)
Missing	1 (0.0)
Race (self-identified)	
European American	142,961 (82.4)
African American	18,238 (10.5)
Asian	1,123 (0.6)
American Indian/Native Hawaiian	1,120 (0.6)
Pacific Islander	205 (0.1)
Other	3,012 (1.7)
Multiple races	6,516 (3.8)
Missing	399 (0.2)
Ethnicity (self-identified)	
Hispanic	10,598 (6.1)
Non-Hispanic	162,186 (93.4)
Unknown	788 (0.5)
Missing	2 (0.0)
Marital status	
Married	104,459 (60.2)
Civil commitment/cohabitating	4,896 (2.8)
Separated	3,630 (2.1)
Divorced	32,722 (18.9)
Widowed	12,594 (7.3)
Never married	13,323 (7.7)
Missing	1,950 (1.1)
Military service-related characteristics	
Military branch	
Army	75,202 (43.3)
Navy	34,627 (19.9)
Air Force	28,028 (16.1)
Marine Corps	16,314 (9.4)
Other	2,447 (1.4)
Multiple	15,402 (8.9)
Missing	1,554 (0.9)

(Continues)

TABLE 1 (Continued)

Factor	Frequency (%) or mean (SD)
Service era	
September 2001 or later	3,530 (2.0)
August 1990 to August 2001	4,008 (2.3)
May 1975 to July 1990	13,255 (7.6)
August 1964 to April 1975	68,867 (39.7)
February 1955 to July 1964	15,401 (8.9)
July 1950 to January 1955	12,490 (7.2)
January 1947 to June 1950	574 (0.3)
December 1941 to December 1946	6,963 (4.0)
November 1941 or earlier	154 (0.1)
Multiple	48,320 (27.8)
Missing	12 (0.0)
PCL score, mean (SD)	30.9 (15.8)
PCL re-experiencing cluster score, mean (SD)	8.5 (4.9)
Deployment and combat related	
Deployed	
Yes	107,208 (61.8)
No	56,867 (32.8)
Missing	9,499 (5.5)
Served in combat or war zone	
Yes	83,097 (47.9)
No	88,228 (50.8)
Missing	2,249 (1.3)
Modified DRRI combat experiences score, mean (SD)	27.4 (11.1)

Note. PCL: Posttraumatic Stress Disorder Checklist; DRRI: Deployment Risk and Resilience Inventory.

were recruited at various centers to boost enrollment. More than 695,000 veterans have enrolled in MVP, and cleaned genotypes were available for 353,948 participants at the time of current analyses. Table 1 shows interim results describing self-reported demographic and military service-related characteristics of MVP enrollees with completed Baseline and Lifestyle surveys ($N = 173,574$), representing the source population for the proposed study. The most common period of military service was the Vietnam era ($n = 68,867$; 39.7%) and the most common branch of service the Army ($n = 75,202$; 43.3%). Deployment was endorsed by 61.8% ($n = 107,208$) of participants and combat exposure by 47.9% ($n = 83,097$).

Other service-related measures in Table 1 include the modified version of the original Deployment Risk and Resilience Inventory (DRRI) combat experiences scale (Vogt, Proctor, King, King, & Vasterling, 2008) and the DSM-IV PCL score. The modified DRRI combat experiences scale, consisting of 15 questions, measures intensity of combat exposure, scored using a 5-point Likert-type response format. The PCL, which assesses severity of PTSD symptoms, contains 17 questions (also scored with the 5-point Likert-type response format), of which the first five together constitute the re-experiencing symptom cluster subphenotype (Forbes et al., 2001), which differentiates PTSD from other mental disorders (Sareen, 2014). The mean and standard deviation (SD) DRRI, PCL, and re-experiencing scores were 27.4 (11.1), 30.9 (15.8), and 8.5 (4.9),

respectively, in the MVP participants. Among combat-exposed veterans, the mean (SD) PCL and re-experiencing scores were higher at 34.2 (17.1) and 9.6 (5.3), respectively.

Self-reported personal habits frequently observed in PTSD patients—smoking, alcohol abuse, reduced sleep hours, and risk taking—are shown in Table 2, along with prevalence of self-reported mental health conditions that commonly co-occur with PTSD. For example, in a typical month, 35.1% ($n = 55,406$) of the respondents may have had five or more alcoholic drinks in 1 day, the amount defined for men as binge drinking by the Substance Abuse and Mental Health Services Administration and as at risk for developing alcohol use disorder by the National Institute on Alcohol Abuse and Alcoholism (2017). Among mental health diagnoses, PTSD was self-reported by 16.2% ($n = 28,141$) of MVP participants and 25.9% ($n = 21,510$) of combat-exposed veterans. Depression was the most commonly endorsed disorder at 25.4% ($n = 44,034$); panic disorder was reported by 14.3% ($n = 24,749$) and memory loss by 9.3% ($n = 16,155$).

3.2 | Optimizing threshold probability for case/control definition

To minimize false-positive and false-negative attributions, the algorithm for automating selection of cases and controls used a

TABLE 2 Self-reported behavioral and health characteristics of genotyped Million Veteran Program enrollees with completed Base-line and Lifestyle surveys ($N = 173,574$)

Factor	Frequency (%)
Personal habits	
Current smoker	
Yes	20,369 (11.7)
No	94,837 (54.6)
Missing	58,368 (33.6)
Maximum no. of alcoholic drinks in 1 day	
0	9,662 (5.6)
1	20,978 (12.1)
2	31,045 (17.9)
3	23,102 (13.3)
4	17,666 (10.2)
5–6	22,318 (12.9)
7–9	10,844 (6.2)
10–14	8,957 (5.2)
≥ 15	13,287 (7.7)
Missing	15,715 (9.1)
Number of hours of sleep per day	
≤ 5	23,832 (13.7)
6	37,329 (21.5)
7	48,324 (27.8)
8	41,984 (24.2)
9	14,012 (8.1)
≥ 10	6,556 (3.8)
Missing	1,537 (0.9)
Risk taking (seat belt use)	
Front seat belt	
Almost always	161,613 (93.1)
Sometimes	7,736 (4.5)
Never	1,995 (1.1)
Not applicable	595 (0.3)
Missing	1,635 (0.9)
Back seat belt	
Almost always	118,768 (68.4)
Sometimes	29,813 (17.2)
Never	18,564 (10.7)
No seat belt	2,061 (1.2)
Not applicable	1,585 (0.9)
Missing	2,783 (1.6)
Self-reported mental health conditions	
Anxiety reaction/panic disorder	24,749 (14.3)
ADHD (attention deficit hyperactivity disorder)	4,201 (2.4)
Bipolar disorder	5,386 (3.1)
PTSD (posttraumatic stress disorder)	28,141 (16.2)
Depression	44,034 (25.4)
Eating disorder	4,050 (2.3)
Personality disorder	4,181 (2.4)
Schizophrenia	2,113 (1.2)
Social phobia	3,936 (2.3)
Other mental health disorder	5,492 (3.2)

(Continues)

TABLE 2 (Continued)

Factor	Frequency (%)
Traumatic brain injury	4,855 (2.8)
Memory loss or impairment	16,155 (9.3)
Dementia	1,925 (1.1)

probabilistic approach, which generates probabilities, related to trait severity, of membership in each of three classes. The threshold probability, $\pi_{\min}$ ($=\pi_{\text{opt}}$), selected for classification as a definite case or control maximized the statistical power to detect association between SNPs and disease. As shown in Table 3, for a disease prevalence of 30% as cited in the study protocol, various combinations of $\pi_{\min}$, MAF, and RR of disease for an allele were simulated. A $\pi_{\min}$ of 0.7 for cases and controls maximized power, with results largely insensitive to MAF and RR. Based on comparisons of telephone interview results with algorithm predictions using these threshold probabilities, refinements were made to the model, including imposition of $\text{PCL} \leq 30$ for predicted controls to be classified as such, to minimize false negatives.

3.3 | Algorithm performance

The performance of the algorithm can be evaluated by comparing its predictions with results of 500 chart reviews conducted by five experts at three different VA centers; see Table 4. Both the algorithm and chart reviewers classified participants into three groups: likely PTSD (case), possible PTSD, and likely not PTSD (control); insufficient information was available in 15 of the 500 charts. The overall percent agreement between reviewer and algorithm was 65.2% ($=316/485$), and weighted $\kappa = 0.62$ (95% confidence interval [CI] [0.57, 0.67]), indicating substantial agreement. The corresponding sensitivity, specificity, positive predictive value, and negative predictive value for combat-exposed participants were 0.78, 0.98, 0.85, and 0.96, respectively (Harrington et al., in press). We focus attention in this work on $\pi_{\min}$ ($=\pi_{\text{opt}}$) = 0.7, given that this threshold maximized statistical power to detect association between SNPs and disease; Harrington et al. (in press) report on how $\pi_{\min}$ affected the sample sizes, estimated misclassification rates, and operating characteristics.

Preliminary comparisons of algorithm predictions with telephone interview results ($n = 132$) showed an overall percent agreement of 90.2% ($=119/132$) and $\kappa = 0.75$ (95% CI [0.62, 0.88]), again indicating substantial agreement. Chart reviews had been conducted on 25 of these 132 participants. The overall percent agreement between chart reviews (primarily DSM-IV criteria) and telephone interviews (DSM-5 criteria) for the 25 veterans was 84.0% ($=21/25$) and $\kappa = 0.68$ (95% CI [0.40, 0.96]), indicating substantial agreement.

The probabilistic approach thus resulted in reasonably accurate predictions, while providing flexibility for performing sensitivity analysis by varying the threshold probability that a deterministic algorithm cannot.

TABLE 3 Statistical power (%) for combinations of minor allele frequency (MAF), relative risk (RR) of disease, and case/control threshold probability ($\pi_{\min}$), for disease prevalence = 30%

Statistical power (%) as a function of MAF and RR				
MAF	RR = 1.05	RR = 1.10	RR = 1.15	RR = 1.20
$\pi_{\min} = 0.60$				
0.1	0.1	17	90	100
0.2	0.6	68	100	100
0.3	1.9	89	100	100
0.5	3.5	95	100	100
$\pi_{\min} = 0.65$				
0.1	0.1	15	88	100
0.2	0.6	64	100	100
0.3	1.6	87	100	100
0.5	3.0	94	100	100
$\pi_{\min} = 0.70$				
0.1	0.1	17	91	100
0.2	0.7	69	100	100
0.3	2.0	90	100	100
0.5	3.8	96	100	100
$\pi_{\min} = 0.75$				
0.1	0.1	17	90	100
0.2	0.7	68	100	100
0.3	1.9	89	100	100
0.5	3.6	95	100	100

TABLE 4 Patient classification as likely posttraumatic stress disorder (PTSD), possible PTSD, and likely not PTSD, comparing the algorithm with chart review as the reference standard

		Chart review (reference standard)			
		Likely PTSD	Possible PTSD	Likely not PTSD	Total
Algorithm	Likely PTSD	117	8	0	125
	Possible PTSD	80	57	61	198
	Likely not PTSD	1	19	142	162
	Total	198	84	203	485

Note. Weighted kappa = 0.62 (95% confidence interval [0.57, 0.67]); 15 participants had insufficient information.

By applying the algorithm to the 74,901 combat-exposed, genotyped MVP participants without a diagnosis of schizophrenia and bipolar disorder, and after discarding possible cases (i.e., $\text{Pr}(\text{case}) \leq 0.7$ and $\text{Pr}(\text{control}) \leq 0.7$) and imposing the $\text{PCL} \leq 30$ criterion for predicted controls, we identified 16,490 PTSD cases (i.e., $\text{Pr}(\text{case}) > 0.7$) and 33,609 controls (i.e., $\text{Pr}(\text{control}) > 0.7$).

4 | DISCUSSION

The high prevalence of military veterans with combat exposure (see Table 1), which is a strong risk factor for PTSD and other comorbidities including suicidality, provides motivation for the current study. Only a modest number (and mostly unreplicated) PTSD risk loci have been identified to date (Almli et al., 2015; Ashley-Koch et al., 2015; Banerjee et al., 2017; Duncan et al., 2018; Nievergelt et al., 2018; Sheerin et al., 2017; Stein et al., 2016). Our MVP-based combat-exposed, genotyped sample, with >16,000 PTSD cases and >33,000

non-PTSD controls, provides a strong foundation to conduct a GWAS. Additional information can be derived from the noncombat-exposed subjects, some of whom have PTSD attributable to other stressors.

In addition to uncertainties associated with previous genomic studies, the definition of PTSD itself has evolved recently (DiMauro, Carter, Folk, & Kashdan, 2014; Galatzer-Levy & Bryant, 2013; Kirkbride, 2012). In particular, recent changes in PTSD criteria from DSM-IV to DSM-5 (Calhoun et al., 2012; Weathers et al., 2018) may represent a challenge, inasmuch as the algorithm and chart review activities were primarily informed by DSM-IV criteria, whereas the telephone interview uses CAPS-5, which is based on DSM-5 but “downward compatible” (Weathers et al., 2018). How the reorganization of DSM-IV symptom clusters in DSM-5 will affect PTSD case/control status is therefore of interest. Although differences between the two sets of criteria have been reported (McFarlane, 2014), several studies noted that DSM-5 identified largely the same cohort of patients as did DSM-IV (Calhoun et al., 2012; Weathers et al., 2018; Wisco et al., 2016). In addition, for a complex trait GWAS,

a large sample size (as with MVP) can quench relatively minor differences in phenotypic definition, such as the difference between DSM-IV and DSM-5 PTSD diagnoses. This situation has been observed for related conditions, such as major depressive disorder (Wray et al., 2018). In the current context, accurate DSM-IV and DSM-5 PTSD are much closer to each other than the differences between either of these and non-PTSD. Albeit limited by small sample size, the agreement between our chart review results using primarily DSM-IV criteria and our telephone interview results using DSM-5 criteria is encouraging.

A related methodological issue is the challenge of using data from EHRs for phenotyping (Ouyang et al., 2016; Sinnott et al., 2014). Accuracy of data in these records is variable, and extracting information can be challenging. To automate selection of cases and controls, we adopted a probabilistic methodology to minimize both false-negative and false-positive designations. An algorithm was constructed and refined based on EHR chart reviews and telephone interviews, with substantial agreement noted between algorithm predictions and results from both chart reviews and telephone interviews. For the primary analysis, which will classify PTSD as a binary case/control trait, the algorithm identified adequate numbers of cases and controls, despite eliminating the possible cases (given the large size of the MVP study population). By eliminating the possible cases, we are more confident in the predicted case/control status of each study participant.

The probabilistic approach of defining PTSD caseness as a continuous (as opposed to a discrete) outcome permits sensitivity analyses to refine further the definition of PTSD. Based on the number of positive PTSD symptoms or probability of PTSD caseness, the condition can also be considered as a quantitative trait. This approach, which has been used in substance use disorder studies (Gelernter et al., 2015), enhances statistical power because it assigns greater weight to the severely affected than to the moderately affected or to those without an unequivocal diagnosis.

In conclusion, the probabilistic approach has identified large samples of combat-exposed cases and controls, with potential for the ongoing GWAS to identify genetic markers to aid the etiological understanding and treatment of PTSD.

DECLARATION OF INTEREST STATEMENT

Dr. Stein has in the past 3 years been a consultant for Actelion, Alkermes, Aptinyx, Dart Neuroscience, Healthcare Management Technologies, Janssen, Neurocrine Biosciences, Oxeia Biopharmaceuticals, Pfizer, and Resilience Therapeutics. Dr. Stein owns founders shares and stock options in Resilience Therapeutics and has stock options in Oxeia Biopharmaceuticals. No other authors report any conflict of interest.

ACKNOWLEDGEMENTS

CSP#575B Planning Committee: M. Aslan, C. Brandt, J. Concato, J. M. Gaziano, J. Gelernter, T. Gleason, K. Koenen, C. Marx, N. Schork, M. Stein, J. Turner, H. Zhao.

CSP#575B Executive Committee: C. Brandt, J. Concato, J. M. Gaziano, J. Gelernter, T. Gleason, G. Huang, K. Koenen, C. Marx, J. Moser, K. Radhakrishnan, N. Schork, M. Stein, H. Zhao.

CSP#575B Study Chairs' Offices: VA Connecticut Healthcare System, West Haven, CT, J. Gelernter (Study Co-Chair), J. Kaufman, Y. Nunez, R. Pietrzak. VA San Diego Healthcare System, San Diego, CA, M. Stein (Study Co-Chair), D. Beck, M. R. Behrooznia, S. Cissell, J. A. Nance, T. R. Smith.

CSP Epidemiology Centers: VA Clinical Epidemiology Research Center (CERC), VA Connecticut Healthcare System, West Haven, CT, J. Concato (Director), M. Aslan, Q. Chen, K-H Cheung, P. Crutchfield, W. Lance, Y. Li, K. Radhakrishnan, N. Rajeevan, F. Sayward, N. Sun; the Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA, J. M. Gaziano (Director), K. Cho, D. Gagnon, K. Harrington, J. Honerlaw, R. Quaden, D. Pratt, S. Pyarajan, S. Whitbourne.

VA Office of Research and Development: T. Gleason (Clinical Science Research and Development Service), G. Huang (Cooperative Studies Program), J. Moser (Program Manager, Million Veteran Program), S. Muralidhar (Million Veteran Program), T. O'Leary and R. Ramoni (Chief Research and Development Officers).

The authors thank the reviewers for helpful comments.

This research is funded by the Department of Veterans Affairs Cooperative Studies Program.

The contents do not represent the views of the U.S. Department of Veterans Affairs or the United States Government.

ORCID

Krishnan Radhakrishnan <https://orcid.org/0000-0001-9243-7733>

REFERENCES

- Almli, L. M., Stevens, J. S., Smith, A. K., Kilaru, V., Meng, Q., Flory, J., ... Ressler, K. J. (2015). A genome-wide identified risk variant for PTSD is a methylation quantitative trait locus and confers decreased cortical activation to fearful faces. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, 168B(5), 327–336. <https://doi.org/10.1002/ajmg.b.32315>
- Ashley-Koch, A. E., Garrett, M. E., Gibson, J., Liu, Y., Dennis, M. F., Kimbrel, N. A., ... Hauser, M. A. (2015). Genome-wide association study of post-traumatic stress disorder in a cohort of Iraq–Afghanistan era veterans. *Journal of Affective Disorders*, 184, 225–234. <https://doi.org/10.1016/j.jad.2015.03.049>
- Aziz, M. A., & Kenford, S. (2004). Comparability of telephone and face-to-face interviews in assessing patients with posttraumatic stress disorder. *Journal of Psychiatric Practice*, 10(5), 307–313. <https://doi.org/10.1097/00131746-200409000-00004>
- Ballard, D., Abraham, C., Cho, J., & Zhao, H. (2010). Pathway analysis comparison using Crohn's disease genome wide association studies. *BMC Medical Genomics*, 3, 25. <https://doi.org/10.1186/1755-8794-3-25>
- Banerjee, S. B., Morrison, F. G., & Ressler, K. J. (2017). Genetic approaches for the study of PTSD: Advances and challenges. *Neuroscience Letters*, 649, 139–146. <https://doi.org/10.1016/j.neulet.2017.02.058>
- Brown, S. H., Lincoln, M. J., Groen, P. J., & Kolodner, R. M. (2003). VistA—U.S. Department of Veterans Affairs national-scale HIS. *International Journal of Medical Informatics*, 69(2–3), 135–156. [https://doi.org/10.1016/S1386-5056\(02\)00131-4](https://doi.org/10.1016/S1386-5056(02)00131-4)
- Calhoun, P. S., Hertzberg, J. S., Kirby, A. C., Dennis, M. F., Hair, L. P., Dedert, E. A., & Beckham, J. C. (2012). The effect of draft DSM-V criteria on post-traumatic stress disorder prevalence. *Depression and Anxiety*, 29(12), 1032–1042. <https://doi.org/10.1002/da.22012>

- Chen, M., Cho, J., & Zhao, H. (2011). Detecting epistatic SNPs associated with complex diseases via a Bayesian classification tree search method. *Annals of Human Genetics*, 75(1), 112–121. <https://doi.org/10.1111/j.1469-1809.2010.00627.x>
- Daskalakis, N. P., Rijal, C. M., King, C., Huckins, L. M., & Ressler, K. J. (2018). Recent genetics and epigenetics approaches to PTSD. *Current Psychiatry Reports*, 20(5), 30. <https://doi.org/10.1007/s11920-018-0898-7>
- Dickstein, B. D., Weathers, F. W., Angkaw, A. C., Nievergelt, C. M., Yurgil, K., Nash, W. P., ... Marine Resiliency Study Team (2015). Diagnostic utility of the Posttraumatic Stress Disorder (PTSD) Checklist for identifying full and partial PTSD in active-duty military. *Assessment*, 22(3), 289–297. <https://doi.org/10.1177/1073191114548683>
- DiMauro, J., Carter, S., Folk, J. B., & Kashdan, T. B. (2014). A historical review of trauma-related diagnoses to reconsider the heterogeneity of PTSD. *Journal of Anxiety Disorders*, 28(8), 774–786. <https://doi.org/10.1016/j.janxdis.2014.09.002>
- Duncan, L. E., Ratanatharathorn, A., Aiello, A. E., Almli, L. M., Amstadter, A. B., Ashley-Koch, A. E., ... Koenen, K. C. (2018). Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. *Molecular Psychiatry*, 23(3), 666–673. <https://doi.org/10.1038/mp.2017.77>
- Forbes, D., Creamer, M., & Biddle, D. (2001). The validity of the PTSD checklist as a measure of symptomatic change in combat-related PTSD. *Behaviour Research and Therapy*, 39(8), 977–986. [https://doi.org/10.1016/S0005-7967\(00\)00084-X](https://doi.org/10.1016/S0005-7967(00)00084-X)
- Fulton, J. J., Calhoun, P. S., Wagner, H. R., Schry, A. R., Hair, L. P., Feeling, N., ... Beckham, J. C. (2015). The prevalence of posttraumatic stress disorder in Operation Enduring Freedom/Operation Iraqi Freedom (OEF/OIF) Veterans: A meta-analysis. *Journal of Anxiety Disorders*, 31, 98–107. <https://doi.org/10.1016/j.janxdis.2015.02.003>
- Galatzer-Levy, I. R., & Bryant, R. A. (2013). 636,120 ways to have posttraumatic stress disorder. *Perspectives on Psychological Science*, 8(6), 651–662. <https://doi.org/10.1177/1745691613504115>
- Galinsky, K. J., Bhatia, G., Loh, P. R., Georgiev, S., Mukherjee, S., Patterson, N. J., & Price, A. L. (2016). Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia. *American Journal of Human Genetics*, 98(3), 456–472. <https://doi.org/10.1016/j.ajhg.2015.12.022>
- Gates, M. A., Holowka, D. W., Vasterling, J. J., Keane, T. M., Marx, B. P., & Rosen, R. C. (2012). Posttraumatic stress disorder in veterans and military personnel: Epidemiology, screening, and case recognition. *Psychological Services*, 9(4), 361–382. <https://doi.org/10.1037/a0027649>
- Gaziano, J. M., Concato, J., Brophy, M., Fiore, L., Pyarajan, S., Breeling, J., ... O'Leary, T. J. (2016). Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *Journal of Clinical Epidemiology*, 70, 214–223. <https://doi.org/10.1016/j.jclinepi.2015.09.016>
- Gelernter, J., Kranzler, H. R., Sherva, R., Almasy, L., Herman, A. I., Koesterer, R., ... Farrer, L. A. (2015). Genome-wide association study of nicotine dependence in American populations: Identification of novel risk loci in both African-Americans and European-Americans. *Biological Psychiatry*, 77(5), 493–503. <https://doi.org/10.1016/j.biopsych.2014.08.025>
- Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., ... Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74. <https://doi.org/10.1038/nature15393>
- Goldstein, R. B., Smith, S. M., Chou, S. P., Saha, T. D., Jung, J., Zhang, H., ... Grant, B. F. (2016). The epidemiology of DSM-5 posttraumatic stress disorder in the United States: Results from the National Epidemiologic Survey on Alcohol and Related Conditions-III. *Social Psychiatry and Psychiatric Epidemiology*, 51(8), 1137–1148. <https://doi.org/10.1007/s00127-016-1208-5>
- Gros, D. F., Szafranski, D. D., Brady, K. T., & Back, S. E. (2015). Relations between pain, PTSD symptoms, and substance use in veterans. *Psychiatry*, 78(3), 277–287. <https://doi.org/10.1080/00332747.2015.1069659>
- Guffanti, G., Galea, S., Yan, L., Roberts, A. L., Solovieff, N., Aiello, A. E., ... Koenen, K. C. (2013). Genome-wide association study implicates a novel RNA gene, the lincRNA AC068718.1, as a risk factor for post-traumatic stress disorder in women. *Psychoneuroendocrinology*, 38(12), 3029–3038. <https://doi.org/10.1016/j.psyneuen.2013.08.014>
- Harrington, K. M., Quaden, R., Stein, M. B., Honerlaw, J. P., Cissell, S., Pietrzak, R. H., ... Cho, K. (in press). Validation of an electronic medical record-based algorithm for identifying posttraumatic stress disorder in US Veterans. *Journal of Traumatic Stress*.
- Holowka, D. W., Marx, B. P., Gates, M. A., Litman, H. J., Ranganathan, G., Rosen, R. C., & Keane, T. M. (2014). PTSD diagnostic validity in Veterans Affairs electronic records of Iraq and Afghanistan veterans. *Journal of Consulting and Clinical Psychology*, 82(4), 569–579. <https://doi.org/10.1037/a0036347>
- Hou, L., Sun, N., Mane, S., Sayward, F., Rajeevan, N., Cheung, K. H., ... Zhao, H. (2017). Impact of genotyping errors on statistical power of association tests in genomic analyses: A case study. *Genetic Epidemiology*, 41(2), 152–162. <https://doi.org/10.1002/gepi.22027>
- Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics*, 44(8), 955–959. <https://doi.org/10.1038/ng.2354>
- Huttunen, H., Manninen, T., Kauppi, J.-P., & Tohka, J. (2013). Mind reading with regularized multinomial logistic regression. *Machine Vision and Applications*, 24(6), 1311–1325. <https://doi.org/10.1007/s00138-012-0464-y>
- Jang, K. L., Taylor, S., Stein, M. B., & Yamagata, S. (2007). Trauma exposure and stress response: Exploration of mechanisms of cause and effect. *Twin Research and Human Genetics*, 10(4), 564–572. <https://doi.org/10.1375/twin.10.4.564>
- Jorge, R. E. (2015). Posttraumatic stress disorder. *Continuum (Minneapolis)*, 21(3 Behavioral Neurology and Neuropsychiatry), 789–805. <https://doi.org/10.1212/01.CON.0000466667.20403.b1>
- Kirkbride, J. F. (2012). PTSD: An elusive definition. *Journal of Special Operations Medicine*, 12(2), 42–47.
- Logue, M. W., Baldwin, C., Guffanti, G., Melista, E., Wolf, E. J., Reardon, A. F., ... Miller, M. W. (2013). A genome-wide association study of post-traumatic stress disorder identifies the retinoid-related orphan receptor alpha (RORA) gene as a significant risk locus. *Molecular Psychiatry*, 18(8), 937–942. <https://doi.org/10.1038/mp.2012.113>
- Lu, Q., Powles, R. L., Wang, Q., He, B. J., & Zhao, H. (2016). Integrative tissue-specific functional annotations in the human genome provide novel insights on many complex traits and improve signal prioritization in genome wide association studies. *PLoS Genetics*, 12(4), e1005947. <https://doi.org/10.1371/journal.pgen.1005947>
- Marmar, C. R., Schlenger, W., Henn-Haase, C., Qian, M., Purchia, E., Li, M., ... Kulka, R. A. (2015). Course of posttraumatic stress disorder 40 years after the Vietnam war: Findings from the National Vietnam Veterans Longitudinal Study. *JAMA Psychiatry*, 72(9), 875–881. <https://doi.org/10.1001/jamapsychiatry.2015.0803>
- Mawanda, F., Wallace, R. B., McCoy, K., & Abrams, T. E. (2017). PTSD, psychotropic medication use, and the risk of dementia among US veterans: A retrospective cohort study. *Journal of the American Geriatrics Society*, 65(5), 1043–1050. <https://doi.org/10.1111/jgs.14756>
- McFarlane, A. C. (2014). PTSD and DSM-5: Unintended consequences of change. *Lancet Psychiatry*, 1(4), 246–247. [https://doi.org/10.1016/S2215-0366\(14\)70321-9](https://doi.org/10.1016/S2215-0366(14)70321-9)
- Murphy, D., Iversen, A., & Greenberg, N. (2008). The mental health of veterans. *Journal of the Royal Army Medical Corps*, 154(2), 136–139. <https://doi.org/10.1136/jramc-154-02-13>
- National Institute on Alcohol Abuse and Alcoholism. (2017). Drinking levels defined. Retrieved from <https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking>
- Nievergelt, C. M., Ashley-Koch, A. E., Dalvie, S., Hauser, M. A., Morey, R. A., Smith, A. K., & Uddin, M. (2018). Genomic approaches to post-traumatic stress disorder: The psychiatric genomic consortium initiative. *Biological Psychiatry*, 83(10), 831–839. <https://doi.org/10.1016/j.biopsych.2018.01.020>

- Nievergelt, C. M., Maihofer, A. X., Mustapic, M., Yurgil, K. A., Schork, N. J., Miller, M. W., ... Baker, D. G. (2015). Genomic predictors of combat stress vulnerability and resilience in U.S. Marines: A genome-wide association study across multiple ancestries implicates PRTFDC1 as a potential PTSD gene. *Psychoneuroendocrinology*, 51, 459–471. <https://doi.org/10.1016/j.psyneuen.2014.10.017>
- Ouyang, L., Apley, D. W., & Mehrotra, S. (2016). A design of experiments approach to validation sampling for logistic regression modeling with error-prone medical records. *Journal of the American Medical Informatics Association*, 23(e1), e71–e78. <https://doi.org/10.1093/jamia/ocv132>
- Phillips, C. J., Leardmann, C. A., Gumbs, G. R., & Smith, B. (2010). Risk factors for posttraumatic stress disorder among deployed US male marines. *BMC Psychiatry*, 10, 52. <https://doi.org/10.1186/1471-244X-10-52>
- Pitman, R. K., Gilbertson, M. W., Gurvits, T. V., May, F. S., Lasko, N. B., Metzger, L. J., ... Orr, S. P. (2006). Clarifying the origin of biological abnormalities in PTSD through the study of identical twins discordant for combat exposure. *Annals of the New York Academy of Sciences*, 1071, 242–254. <https://doi.org/10.1196/annals.1364.019>
- Polimanti, R., Kaufman, J., Zhao, H., Kranzler, H. R., Ursano, R. J., Kessler, R. C., ... Stein, M. B. (2018). A genome-wide gene-by-trauma interaction study of alcohol misuse in two independent cohorts identifies PRKG1 as a risk locus. *Molecular Psychiatry*, 23(1), 154–160. <https://doi.org/10.1038/mp.2017.24>
- Prescott, M. R., Tamburrino, M., Calabrese, J. R., Liberzon, I., Slembariski, R., Shirley, E., ... Galea, S. (2014). Validation of lay-administered mental health assessments in a large Army National Guard cohort. *International Journal of Methods in Psychiatric Research*, 23(1), 109–119. <https://doi.org/10.1002/mpr.1416>
- Rodgers, A. B., & Bale, T. L. (2015). Germ cell origins of posttraumatic stress disorder risk: The transgenerational impact of parental stress experience. *Biological Psychiatry*, 78(5), 307–314. <https://doi.org/10.1016/j.biopsych.2015.03.018>
- Sareen, J. (2014). Posttraumatic stress disorder in adults: Impact, comorbidity, risk factors, and treatment. *Canadian Journal of Psychiatry*, 59(9), 460–467. <https://doi.org/10.1177/070674371405900902>
- Sheerin, C. M., Lind, M. J., Bountress, K. E., Nugent, N. R., & Amstadter, A. B. (2017). The genetics and epigenetics of PTSD: Overview, recent advances, and future directions. *Current Opinion in Psychology*, 14, 5–11. <https://doi.org/10.1016/j.copsyc.2016.09.003>
- Sinnott, J. A., Dai, W., Liao, K. P., Shaw, S. Y., Ananthakrishnan, A. N., Gainer, V. S., ... Cai, T. (2014). Improving the power of genetic association tests with imperfect phenotype derived from electronic medical records. *Human Genetics*, 133(11), 1369–1382. <https://doi.org/10.1007/s00439-014-1466-9>
- Skol, A. D., Scott, L. J., Abecasis, G. R., & Boehnke, M. (2006). Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature Genetics*, 38(2), 209–213. <https://doi.org/10.1038/ng1706>
- Smoller, J. W. (2016). The genetics of stress-related disorders: PTSD, depression, and anxiety disorders. *Neuropsychopharmacology*, 41(1), 297–319. <https://doi.org/10.1038/npp.2015.266>
- Stein, M. B. (2018). Genomics of posttraumatic stress disorder: Sequencing stress and modeling misfortune. *Biological Psychiatry*, 83(10), 795–796. <https://doi.org/10.1016/j.biopsych.2017.05.001>
- Stein, M. B., Campbell-Sills, L., Gelernter, J., He, F., Heeringa, S. G., Nock, M. K., ... Army STARRS Collaborators (2017). Alcohol misuse and co-occurring mental disorders among new soldiers in the U.S. Army. *Alcoholism, Clinical and Experimental Research*, 41(1), 139–148. <https://doi.org/10.1111/acer.13269>
- Stein, M. B., Chen, C. Y., Ursano, R. J., Cai, T., Gelernter, J., Heeringa, S. G., ... Smoller, J. W. (2016). Genome-wide association studies of posttraumatic stress disorder in 2 cohorts of US Army soldiers. *JAMA Psychiatry*, 73(7), 695–704. <https://doi.org/10.1001/jamapsychiatry.2016.0350>
- Stein, M. B., & Smoller, J. W. (2018). Precision psychiatry—Will genomic medicine lead the way? *JAMA Psychiatry*, 75(7), 663–664. <https://doi.org/10.1001/jamapsychiatry.2018.0375>
- Sullivan, P. F., Agrawal, A., Bulik, C. M., Andreassen, O. A., Borglum, A. D., Breen, G., ... O'Donovan, M. C. (2018). Psychiatric genomics: An update and an agenda. *The American Journal of Psychiatry*, 175(1), 15–27. <https://doi.org/10.1176/appi.ajp.2017.17030283>
- Vogt, D. S., Proctor, S. P., King, D. W., King, L. A., & Vasterling, J. J. (2008). Validation of scales from the Deployment Risk and Resilience Inventory in a sample of Operation Iraqi Freedom veterans. *Assessment*, 15(4), 391–403. <https://doi.org/10.1177/1073191108316030>
- Weathers, F. W., Bovin, M. J., Lee, D. J., Sloan, D. M., Schnurr, P. P., Kaloupek, D. G., ... Marx, B. P. (2018). The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5): Development and initial psychometric evaluation in military veterans. *Psychological Assessment*, 30(3), 383–395. <https://doi.org/10.1037/pas0000486>
- Wisco, B. E., Marx, B. P., Miller, M. W., Wolf, E. J., Mota, N. P., Krystal, J. H., ... Pietrzak, R. H. (2016). Probable posttraumatic stress disorder in the US veteran population according to DSM-5: Results from the National Health and Resilience in Veterans Study. *The Journal of Clinical Psychiatry*, 77(11), 1503–1510. <https://doi.org/10.4088/JCP.15m10188>
- Wisco, B. E., Marx, B. P., Wolf, E. J., Miller, M. W., Southwick, S. M., & Pietrzak, R. H. (2014). Posttraumatic stress disorder in the US veteran population: Results from the National Health and Resilience in Veterans Study. *The Journal of Clinical Psychiatry*, 75(12), 1338–1346. <https://doi.org/10.4088/JCP.14m09328>
- Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., ... Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, 50(5), 668–681. <https://doi.org/10.1038/s41588-018-0090-3>
- Wu, T. T., Chen, Y. F., Hastie, T., Sobel, E., & Lange, K. (2009). Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics*, 25(6), 714–721. <https://doi.org/10.1093/bioinformatics/btp041>
- Xie, P., Kranzler, H. R., Yang, C., Zhao, H., Farrer, L. A., & Gelernter, J. (2013). Genome-wide association study identifies new susceptibility loci for posttraumatic stress disorder. *Biological Psychiatry*, 74(9), 656–663. <https://doi.org/10.1016/j.biopsych.2013.04.013>
- Zamorski, M. A., & Boulous, D. (2014). The impact of the military mission in Afghanistan on mental health in the Canadian Armed Forces: A summary of research findings. *European Journal of Psychotraumatology*, 5. <https://doi.org/10.3402/ejpt.v5.23822>
- Zhan, X., Hu, Y., Li, B., Abecasis, G. R., & Liu, D. J. (2016). RVTESTS: An efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics*, 32(9), 1423–1426. <https://doi.org/10.1093/bioinformatics/btw079>

How to cite this article: Radhakrishnan K, Aslan M, Harrington KM, et al. Genomics of posttraumatic stress disorder in veterans: Methods and rationale for Veterans Affairs Cooperative Study #575B. *Int J Methods Psychiatr Res*. 2019;28:e1767. <https://doi.org/10.1002/mpr.1767>