

## Supplementary Online Content

Haley RW, Charuvastra E, Shell WE, et al. Cholinergic autonomic dysfunction among veterans with Gulf War syndrome: confirmation in a population-based sample. *JAMA Neurol*. Published online November 21, 2012. doi:10.1001/jamaneurol.2013.596.

**eAppendix.** Detailed methods.

This supplementary material has been provided by the authors to give readers additional information about their work.

## eAPPENDIX. DETAILED METHODS

### STUDY DESIGN

We performed a nested case-control study comparing clinical and neuroimaging characteristics of chronically ill veterans who were deployed to the Kuwaiti Theater of Operations in the 1991 Persian Gulf War and who met a case definition of Gulf War illness (the cases) with healthy veterans not meeting the case definition who were also deployed (deployed controls) or who were in the U.S. military in 1991 but were not deployed (nondeployed controls). The cases and controls were selected by random sampling from strata of a sampling frame drawn from a national survey of a population-representative sample of the 1991 U.S. military population,<sup>1</sup> explained below.

### CASE DEFINITION

Originally developed by principle components factor analysis of symptoms measured in a 1994-1995 epidemiologic survey of Gulf War veterans from a U.S. Naval Reserve construction battalion, the case definition is comprised of 6 syndrome variants, 3 of which are primary and the other 3 highly overlap with syndrome variant 2.<sup>2</sup> This study focused on the 3 primary syndromes: syndrome 1 (“impaired cognition”), syndrome 2 (“confusion-ataxia”) and syndrome 3 (“central neuropathic pain”). After its original development in members of the Naval Seabees battalion,<sup>2</sup> the case definition was validated with confirmatory factor analysis first in a sample of primarily U.S. Army Gulf War veterans in a VA medical center’s Gulf War illness clinic<sup>3</sup> and then in the national population-representative sample surveyed described below.<sup>1</sup> In a series of clinical case-control studies, the syndrome groups have been shown to differ importantly from each other and from non-ill veterans on war-related environmental exposures,<sup>4</sup> neuropsychological tests,<sup>5,6</sup> neurophysiologic parameters,<sup>5,7</sup> brain metabolism measured by magnetic resonance spectroscopy,<sup>8</sup> functional electroencephalographic (EEG) tests,<sup>9</sup> and response of regional cerebral blood flow, measured by SPECT or arterial spin labeling MRI, to a cholinergic challenge.<sup>10,11</sup>

### SAMPLING FRAME

The sample for this study was drawn from the Gulf War-era U.S. military population in 3 stages. **Stage 1**, performed from 2007 to 2010, involved a computer-assisted telephone interview (CATI) survey of a population-representative sample (N = 8020) drawn from the computerized personnel file of the 1991 U.S. military population (Defense Manpower Data Center, Seaside, CA), stratified by deployment to the Kuwaiti Theater of Operations and other characteristics (**Figure 1**). The methods and findings of this national population survey, known as the U.S. Military Health Survey (USMHS), were published previously.<sup>1</sup> A computerized algorithm of the case definition, applied during the telephone interview, identified all participants meeting the criteria for 1 or more of the 6 syndromes as well as those meeting the Centers for Disease Control and Prevention (CDC) multisymptom illness case definition<sup>12</sup> and the modified Kansas case definition.<sup>13</sup>

The **stage 2** sample was selected at the end of each telephone interview when the interviewer invited a subsample of interview participants (N = 2100) to participate in a nested

case-control study, the Study of Genetic Susceptibility to Gulf War Illness, by donating a blood sample to be analyzed along with the survey information (**Figure 1**). Those invited included all respondents meeting any of the case definitions (cases), random samples of those scoring just below the cutoff for the syndrome definitions (subsyndromic), and those clearly not meeting the definitions (controls). Preliminary findings of the resulting stage 2 nested case-control study were recently presented.<sup>14</sup>

For the present Clinical and Neuroimaging Study, we selected a **stage 3** nested case-control subsample especially suited for an intensive clinical evaluation involving autonomic nervous system testing and extensive magnetic resonance imaging (MRI) of the brain. The *sampling frame* for selecting the stage 3 sample of cases and controls included all of the 2,100 participants the first nested case-control sample who were 65 years of age or younger; weighing 240 pounds or less and having a body mass index under 34 (to fit in the MRI scanner); having no history of claustrophobia, a learning disability or diabetes; and having no history of heavy alcohol dependence or sustained alcohol abuse (**Figure 1**). Information on other comorbid conditions was collected during the visit to be controlled for in the analyses.

The sampling frame was stratified by age (<50, ≥50); a 4-category combination of military rank, gender and race (male, non-black and enlisted; male, black and enlisted; female enlisted; officer); and 5 clinical groups (non-deployed controls; deployed controls; and deployed veterans in the Gulf War illness syndrome 1, syndrome 2, and syndrome 3 groups), thereby producing 8 strata within each of the 5 clinical groups. The deployed and nondeployed control groups consisted of veterans who did not meet the criteria for any of the 6 syndromes in the original case definition<sup>1,2</sup> or the CDC or the modified-Kansas case definitions.<sup>12,13</sup>

#### SAMPLE SIZE TARGET

The goal was to select veterans randomly from the strata of the stage 3 sampling frame to obtain a sample size of approximately 20 deployed veterans in each of the 3 primary syndrome groups and 30 controls equally distributed between the non-deployed and the deployed. A sample size of ≥20 per group is the number needed to provide power of 0.8 for functional MRI experiments adjusted for multiple comparisons,<sup>15</sup> a priority of this study. Although from prior studies we expected no difference between the 2 subgroups of controls, we provided sufficient numbers to test this before combining them into a single control group.

#### SUBJECT RECRUITMENT AND PARTICIPATION

A total of 587 veterans were in the 8 strata of the stage 3 sampling frame. We sorted those in each stratum by a random number and selected the first-listed subjects in each stratum to contact to meet the sample size target (**Figure 1**). To replace attrition from inability to contact or ineligibility, we selected the next subjects on the randomly ordered list in each stratum until the desired final sample size was obtained. Ultimately 235 veterans were selected to be contacted (**Figure 1**). Despite use of publicly available locating databases and telephoning at all hours including evenings and weekends, we were unable to locate 50 of the veterans, who fell disproportionately in the more functionally impaired syndrome 2 and 3 groups (**Figure 1**).

The 185 located subjects were interviewed at length by a physician with extensive experience with ill Gulf War veterans (R.W.H.) who, after obtaining verbal informed consent, took a thorough medical history to confirm the symptoms reported on the CATI survey and the

case definition and detect possible contraindications to participation in MRI scanning. Fifty-nine subjects were found to be ineligible (**Figure 1**). The most common reason for ineligibility (N = 32) was that subjects' verbal description of their symptoms did not confirm the elements of the syndrome case definition that had been elicited earlier in their computer-assisted telephone interview in the USMHS.<sup>1</sup> Among the 169 interviewed subjects meeting the case definition on the survey, the rate of misclassification was strongly related to branch of service during the war, with 15% of cases in Army personnel misclassified, 13% in the Marines, 11% in the Air Force, and 56% in the Navy based on ships in the Persian Gulf ( $P < .001$ ). In the 3 land-based branches, the misclassification rate did not vary significantly across the categories of important exposure measures, such as having heard nerve gas alarms, being in the plume of the Khamisiyah ammunition dump demolition, wearing pesticide-impregnated uniforms, having protective PON1 genotype, and taking pyridostigmine tablets—indicating nondifferential misclassification.

Of the 126 found eligible to participate, 101 agreed and 97 eventually traveled to Dallas for the study (**Figure 1**). The demographic characteristics and comorbidities of the final sample are given in **Table 1**.

## CLINICAL RESEARCH PROTOCOL

All subjects were admitted to the UT Southwestern Clinical and Translational Research Center (CTRC) located in Parkland Memorial Hospital where coffee drinking and smoking were allowed to continue. All subjects gave written informed consent according to a protocol approved by the institutional review boards of our university. An experienced clinical psychologist interviewed all subjects following the Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID)<sup>16</sup> and the Clinician-Administered PTSD Scale (CAPS).<sup>17</sup> All investigators who performed or interpreted tests were blinded to the subjects' case- or control-group status.

On admission to the CTRC the subjects were given the Autonomic Symptom Profile questionnaire<sup>18</sup> to complete. At 07:00 of the second or third hospital day subjects were escorted to the Autonomic Laboratory for a series of non-invasive tests to measure the function of the autonomic nervous system.<sup>19</sup> They had fasted and abstained from alcohol and caffeine for 12 hours overnight, and had discontinued medications with effects on autonomic function (e.g., tricyclic antidepressants, acetylcholinesterase inhibitors) for at least 24 hours prior to testing, but medications taken were recorded for consideration in the analysis. The following procedures were performed on all subjects in the order listed.

*Autonomic Symptoms.* The subjects filled out the Autonomic Symptom Profile,<sup>18</sup> a questionnaire measuring autonomic symptoms that has been validated in normal subjects and patients with autonomic failure. Standard weights were applied to construct the Composite Autonomic Symptom Scale (COMPASS) and the subscales of autonomic symptom domains.<sup>18</sup>

*Peripheral cholinergic autonomic nerves.* The *quantitative sudomotor axon reflex test (QSART)*,<sup>19</sup> which evaluates the function of postganglionic sympathetic cholinergic fibers, was performed at 4 sites on the left side (forearm, proximal lateral leg, medial distal leg, and proximal foot) with the subject lying comfortably on an exam table. As a stimulus, acetylcholine was iontophoresed into the skin using as a 2mA stimulus for 5 minutes, and responses were recorded in a single compartment of a multi-compartmental sweat cell separated from the

stimulus compartment. Total sweat output during the 10 minute recording was recorded at each site.

*Autonomic control of heart rate.* Heart rate responses were monitored with a standard 3-lead electrocardiogram during deep breathing and Valsalva maneuver. Subjects performed deep breathing at 6 times per minute for 8 cycles following a visual prompt, and performed a Valsalva maneuver by blowing into a manometer to generate 40mm Hg for 15 seconds.

*Adrenergic vasomotor function.* The response of blood pressure and heart rate to the Valsalva maneuver and head-up tilt were measured continuously with a Finapres Monitor (Ohmeda, Englewood, Colorado) as the tilt table was brought up to 70 degrees for 10 minutes. The accuracy of beat-to-beat blood pressure recordings were verified by intermittent manual measurement of the blood pressure using an automated sphygmomanometer cuff. If subjects experienced progressive light-headedness or impending syncope during head-up tilt, the table was immediately returned to horizontal.

*Autonomic summary measure.* The Composite Autonomic Severity Score (CASS), a standardized semiquantitative score from 0 (no deficit) to 10 (maximal deficit), was calculated by combining the results of the following 3 subsets of autonomic tests and adjusting to standard age and sex: sudomotor (range, 0-3), cardiovagal (range, 0-3), and adrenergic (range, 0-4).<sup>19</sup> The severity and distribution of autonomic failure is reflected in the total and subset scores.

*Pupillometry.* Autonomic control of pupillary function was measured with a binocular pupillometer (A-1000, Neuroptics Inc, San Clemente, CA), which uses 2 infrared cameras with a digital image capture rate of 30 Hz. Pupil diameter was detected by threshold detection of the dark pupil and corrected for distance from the camera. A light stimulus (calibrated for intensity and duration) was presented to 1 eye using a circumferential array of white light emitting diodes. All data were obtained while subjects were in a comfortable seated position in a darkened room. Dynamic recordings of pupil diameter were saved for offline analysis.

*Lacrimation test.* With a calibrated sterile paper strip (“Color Bar” Schirmer Tear Test strips, Eagle Vision, Memphis, TN) placed inside the lateral lower eyelid, the subjects kept their eyes gently closed for 5 minutes, and the amount of tear production was measured by the migration of moisture zone along the paper strip.

*Circadian heart rate variability.* Autonomic control of cardiovagal function was tested by spectral analysis of a 24-hour Holter recording of the electrocardiogram (EKG).<sup>20-23</sup> Twenty-four hour Holter recordings, performed at home, were digitized at high resolution, and all QRS complexes reviewed on a Pathfinder 710 (Reynolds Medical) by a skilled technician who censored aberrant complexes and artifacts. The normal-to-normal R-R intervals in a 5-minute epoch every 15 minutes were analyzed in the frequency domain using a fast Fourier transform algorithm based on the Lomb-Scargle method of spectral analysis<sup>24</sup> to produce the standard measures of high frequency (HF, 0.15 to <0.40 Hz), low frequency (LF, 0.04 to <0.14 Hz) and very low frequency (VLF, 0.003 to <0.04 Hz) spectral power, expressed in msec<sup>2</sup>. High frequency spectral power of heart rate variability (HF HRV) is an index mainly of vagal parasympathetic influence on cardiac rhythm and is reproducible over time.<sup>25,26</sup>

*Quantitative sensory testing.* Cooling and heat pain thresholds were measured in a quiet, distraction-free room by a single trained technician by non-reaction-time-dependent methods with the CASE IV system (WR Medical Electronics Co., Stillwater, MN).<sup>27</sup>

## STATISTICAL METHODS

The magnitude of the differences among the 4 clinical groups (syndromes 1-3 and controls) on the ASP was estimated by the  $R^2$  statistic from a 4-group analysis of variance. The significance of the group differences on the ASP and the objective autonomic tests comprising the CASS was tested with the Kruskal-Wallis test. The correlation between the ASP symptom domain scales and the CASS autonomic function scales was tested with partial Spearman's rank order correlation, controlling for age, sex and race (black vs other). The interaction of the group effect on sudomotor function with sex described by Stein et al. was tested by the group-by-sex interaction in a 2-way analysis of variance of the QSART test result in the foot.

Hourly measurements of heart-rate variability were log-transformed and analyzed with repeated-measures mixed effects models in which clinical groups (syndromes 1-3 vs controls), time (day vs night) and covariates were treated as fixed effects, and subject as a random effect. Reported mean HF HRV for cases and controls at night and during the day were least square means, and their standard errors, adjusted for the fixed and random effects in the mixed effects models. Night (24:00 to 05:00) and day (08:00 to 21:00) were defined to minimize misclassification by variation in times of sleep onset and waking. In the mixed effects models and ANOVAs Dunnett's correction for multiple comparisons was applied.

Reported results were adjusted for age, sex and race (black vs other), and analyses were rerun to test for confounding by the following covariates: officer rank; smoking; hemoglobin A1c value; indicators of deconditioning BMI and resting pulse rate; creatinine clearance; SCID diagnoses of alcohol or drug abuse and major depressive disorder; CAPS diagnosis of post-traumatic stress disorder (PTSD); diagnosis of heart disease; or medications, including anticholinergics and tricyclic antidepressants, the subjects were taking during the period of the study. Since all the subjects of this nationally representative sample were flown into Dallas for the study, medications could not be discontinued until subjects arrived in the CTSC under medical supervision and thus medications could be discontinued for only 24-48 hours (not for a full 5 half lives) before autonomic testing. Whereas full washout is critical for clinical testing of individual subjects, biasing effects of medications can be assessed and corrected in group comparisons by controlling for them in multivariable models, which were employed.

Statistical tests were performed with the *FREQ*, *CORR*, *NPARIWAY*, *ANOVA* and *MIXED* procedures of SAS for Windows (release 9.2, SAS Institute, Cary, NC). *P* values are 2-tailed.

## REFERENCES

1. Iannacchione VG, Dever JA, Bann CM, et al. Validation of a research case definition of Gulf War illness in the 1991 U.S. military population. *Neuroepidemiol.* 2011;37:129-140.
2. Haley RW, Kurt TL, Hom J. Is there a Gulf War Syndrome? Searching for syndromes by factor analysis of symptoms. *JAMA.* 1997;277:215-222.

3. Haley RW, Luk GD, Petty F. Use of structural equation modeling to test the construct validity of a case definition of Gulf War syndrome: invariance over developmental and validation samples, service branches and publicity. *Psychiatr Res*. 2001;102:175-200.
4. Haley RW, Kurt TL. Self-reported exposure to neurotoxic chemical combinations in the Gulf War. A cross-sectional epidemiologic study. *JAMA*. 1997;277:231-237.
5. Haley RW, Hom J, Roland PS, et al. Evaluation of neurologic function in Gulf War veterans. A blinded case-control study. *JAMA*. 1997;277:223-230.
6. Hom J, Haley RW, Kurt TL. Neuropsychological correlates of Gulf War syndrome. *Arch Clin Neuropsychol*. 1997;12:531-544.
7. Roland PS, Haley RW, Yellin W, Owens K. Vestibular dysfunction in Gulf War syndrome. *Otolaryngol Head Neck Surg*. 2000;122:319-329.
8. Haley RW, Marshall WW, McDonald GG, Daugherty M, Petty F, Fleckenstein JL. Brain abnormalities in Gulf War syndrome: evaluation by <sup>1</sup>H magnetic resonance spectroscopy. *Radiol*. 2000;215:807-817.
9. Tillman GD, Green TA, Ferree TC, et al. Impaired response inhibition in ill Gulf War veterans. *J Neurol Sci*. 2010;297:1-5.
10. Haley RW, Spence JS, Carmack PS, et al. Abnormal brain response to cholinergic challenge in chronic encephalopathy from the 1991 Gulf War. *Psychiatr Res Neuroimag*. 2009;171:207-220.
11. Li X, Spence JS, Buhner DM, et al. Hippocampal dysfunction in Gulf War veterans: investigation with ASL perfusion MR imaging and physostigmine challenge. *Radiol*. 2011;261:218-225.
12. Fukuda K, Nisenbaum R, Stewart G, et al. Chronic multisymptom illness affecting Air Force veterans of the Gulf War. *JAMA*. 1998;280:981-988.
13. Steele L. Prevalence and patterns of Gulf War illness in Kansas veterans: association of symptoms with characteristics of person, place, and time of military service. *Am J Epidemiol*. 2000;152:992-1002.
14. Haley, RW, Kramer, GL, Xiao, J, Teiber, JF. Nerve agent exposure associated with Gulf War encephalopathy through gene-environment interaction with the Q192R polymorphism of *Paraoxonase 1 (PON1)*. 49th Annual Meeting of the Society of Toxicology, Salt Lake City, UT, March 7-11, 2010; Neurotoxicity and Neurodegenerative Disease Session, abstract no. 2242.
15. Desmond JE, Glover GH. Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. *J Neurosci Meth*. 2002;118:115-128.
16. First MB, Spitzer RL, Gibbon M, Williams J. *Structured clinical interview for axis I DSM-IV disorders*. New York: Biometrics Research Department; 1996.
17. Blake DD, Weathers FW, Nagy LM, et al. The development of a clinician-administered PTSD scale. *J Traumatic Stress*. 1995;8:75-90.
18. Suarez GA, Opfer-Gehrking TL, Offord KP, Atkinson EJ, O'Brien PC, Low PA. The Autonomic Symptom Profile: a new instrument to assess autonomic symptoms. *Neurol*. 1999;52:523-528.
19. Low PA. Composite autonomic scoring scale for laboratory quantification of generalized autonomic failure. *Mayo Clin Proc*. 1993;68:748-752.
20. Malik M, Camm AJ. *Heart rate variability*. Armonk, NY: Futura Publishing Company, Inc.; 1995.

21. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. *Circulation*. 1996;93:1043-1065.
22. Berntson GG, Bigger JT, Jr., Eckberg DL, et al. Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiol*. 1997;34:623-648.
23. Eckberg DL. Physiological basis for human autonomic rhythms. *Ann Med*. 2000;32:341-349.
24. Press WH, Rybicki GB. Fast algorithm for spectral analysis of unevenly sampled data. *Astrophysical J*. 1989;338:277-280.
25. Kamalesh M, Burger AJ, Kumar S, Nesto R. Reproducibility of time and frequency domain analysis of heart rate variability in patients with chronic stable angina. *Pacing Clin Electrophysiol*. 1995;18:1991-1994.
26. Burger AJ, Charlamb M, Weinrauch LA, D'Elia JA. Short- and long-term reproducibility of heart rate variability in patients with long-standing type I diabetes mellitus. *Am J Cardiol*. 1997;80:1198-1202.
27. Dyck PJ, O'Brien PC, Kosanke JL, Gillen DA, Karnes JL. A 4, 2, and 1 stepping algorithm for quick and accurate estimation of cutaneous sensation threshold. *Neurol*. 1993;43:1508-1512.