Exercise alters cerebellar and cortical activity related to working memory in phenotypes of Gulf War Illness

Stuart D. Washington,1 Rakib U. Rayhan,2 Richard Garner,1 Destie Provenzano,1 Kristina Zajur,1 Florencia Martinez Addiego,1 John W. VanMeter3 and James N. Baraniuk1
1 Department of Medicine, Georgetown University Medical Center, Washington, DC 20057, USA
2 Department of Physiology and Biophysics, Howard University College of Medicine, Washington, DC 20059, USA
3 Center for Functional and Molecular Imaging, Georgetown University Medical Center, Washington, DC 20057, USA

Gulf War Illness affects 25–32% of veterans from the 1990–91 Persian Gulf War. Post-exertional malaise with cognitive dysfunction, pain and fatigue following physical and/or mental effort is a defining feature of Gulf War Illness. We modelled post-exertional malaise by assessing changes in functional magnetic resonance imaging at 3T during an N-Back working memory task performed prior to a submaximal bicycle stress test and after an identical stress test 24 h later. Serial trends in postural changes in heart rate between supine and standing defined three subgroups of veterans with Gulf War Illness: Postural Orthostatic Tachycardia Syndrome (GWI-POTS, 15%, n = 11), Stress Test Associated Reversible Tachycardia (GWI-START, 31%, n = 23) and Stress Test Originated Phantom Perception (GWI-STOPP, no postural tachycardia, 54%, n = 46). Before exercise, there were no differences in blood oxygenation level-dependent activity during the N-Back task between control (n = 31), GWI-START, GWI-STOPP and GWI-POTS subgroups. Exercise had no effects on blood oxygenation level-dependent activation in controls. GWI-START had post-exertional deactivation of cerebellar dentate nucleus and vermis regions associated with working memory. GWI-STOPP had significant activation of the anterior supplementary motor area that may be a component of the anterior salience network. There was a trend for deactivation of the vermis in GWI-POTS after exercise. These patterns of cognitive dysfunction were apparent in Gulf War Illness only after the exercise stressor. Mechanisms linking the autonomic dysfunction of Stress Test Associated Reversible Tachycardia and Postural Orthostatic Tachycardia Syndrome to cerebellar activation, and Stress Test Originated Phantom Perception to cortical sensorimotor alterations, remain unclear but may open new opportunities for understanding, diagnosing and treating Gulf War Illness.

Correspondence to: Stuart D. Washington, PhD
Department of Medicine
Georgetown University Medical Center
3900 Reservoir Road NW
Washington, DC 20057, USA
E-mail: sdw4@georgetown.edu

Keywords: Gulf War Illness; verbal working memory; cerebellum; supplementary motor area; post-exertional malaise
Abbreviations: AAL = automated anatomical labelling atlas; BOLD = blood oxygenation level dependent; bpm = beats per minute; fMRI = functional magnetic resonance imaging; FoV = field of view; EPI = echo-planar imaging; FWE = family-wise error correction; GWI = Gulf War illness; GWI-START = Gulf War Illness with stress test associated reversible tachycardia; GWI-STO = Gulf War Illness with stress test originated phantom perception; GWI-POTS = Gulf War Illness with postural orthostatic tachycardia syndrome; HC = healthy control; HR = heart rate; HSD = Tukey’s honest significant difference; MNI = Montreal Neurological Institute; ROI = region of interest; SD = standard deviation; SUIT = spatially unbiased infra-tentorial template atlas; T1 = longitudinal relaxation time; TE = echo time; TI = inversion time

Introduction

Gulf War Illness (GWI) affects approximately a quarter of veterans from the 1990–91 Persian Gulf War (Fukuda et al., 1998; Steele, 2000; Research Advisory Committee on Gulf War Veterans’ Illnesses U.S. Department of Veterans Affairs, 2008; Steele et al., 2012). Symptoms include debilitating fatigue, chronic widespread pain, gastrointestinal complaints and cognitive impairment. Diagnosis is based on subjective symptoms of fatigue, mood/cognition and bodily pain using the 1998 Center for Disease Control criteria (Fukuda et al., 1998), and moderate to severe complaints in at least three out of these six categories: fatigue/sleep, neurological/mood/cognition, pain, gastrointestinal, respiratory and skin symptom domains of the 2000 Kansas criteria (Steele, 2000). Although GWI has been attributed to somatoform, depressive and other psychiatric causes (Fiedler et al., 2006; Dursa et al., 2016; Janulewicz et al., 2017), evidence is accumulating for chronic neurotoxic pathologies that were induced by exposures during the Gulf War (White et al., 2016; Georgopoulos et al., 2017). An important clinical finding in GWI is post-exertional malaise with severe symptom exacerbations following physical, cognitive, emotional or other effort (Steele, 2000).

We previously studied the physical and cognitive domains of post-exertional malaise by having subjects perform two submaximal bicycle exercise stress tests on consecutive days with functional magnetic resonance imaging (fMRI) before the first and after the second exercise (Rayhan et al., 2013). We evaluated cognitive dysfunction using the N-Back verbal working memory task during fMRI data acquisition (Owen et al., 2005). Working memory tasks activate frontal-parietal executive control networks. We employed a continuous, verbal version of the N-Back in which subjects performed the simple 0-Back stimulus matching task (‘see the letter, push the button’), and the high cognitive load 2-Back task that involves viewing a series of nine letters at 2 s intervals, remembering their order, then pressing buttons corresponding to the letter seen two places (4 s) previously. Regional brain activation was determined by contrasting the blood oxygenation level-dependent (BOLD) signals of the whole brain between the 2-Back and 0-Back trials in order to identify regions that were differentially activated in this 2-Back > 0-Back condition. Preliminary qualitative analysis found that when all GWI subjects were combined into a single group, there were no differences in BOLD between GWI and control groups, or between pre-exertion and post-exertion MRI studies. However, two patterns of activation and exercise-induced BOLD alterations were found when GWI subjects were stratified by exercise-induced changes in postural orthostatic tachycardia.

In that study, because orthostatic intolerance complaints are common in GWI veterans, we assessed postural changes in heart rate (HR) before exercise and the time course of potential alterations after exercise. Before exercise, GWI veterans had normal elevation of HRs between supine and standing \(DHR = 12 \pm 5\) beats per min (bpm), mean\(\pm\)SD). However, after the first stress test, one-third developed postural tachycardia with \(DHR\) exceeding 30 bpm at least twice during their 5 min periods of standing. This exaggerated orthostatic effect was transient as \(DHR\) returned to normal within 24 to 48 h after exercise (Rayhan et al., 2013; Garner et al., 2018).
The phenomenon was termed Stress Test Activated Reversible Tachycardia (GWI-START). START was distinct from Postural Orthostatic Tachycardia Syndrome (GWI-POTS), which is defined by persistent postural tachycardia with ΔHR ≥30 bpm during virtually all periods of standing (Sheldon et al., 2015). The other two-thirds of GWI subjects did not develop postural tachycardia at any time and were termed the Stress Test Originated Phantom Perception (GWI-STOOP) group because their perceptions of pain were reminiscent of phantom limb pain (Romero-Romo et al., 2010; De Ridder et al., 2014).

Qualitative analysis of BOLD outcomes for the 2-Back > 0-Back condition suggested differences in patterns of brain activation between the GWI-START and GWI-STOOP groups and generated the hypothesis that exertion induced distinct pathological mechanisms of cognitive dysfunction in each subgroup (Rayhan et al., 2013). Here, this hypothesis was tested by recruiting a second wave of healthy control (HC) and GWI subjects to perform the identical exercise-MRI protocol. Quantitative analysis used paired t-tests to identify regions of interest (ROIs) with significantly different pre- and post-exertion BOLD responses for any of the HC, GWI-START, GWI-STOOP or GWI-POTS subgroups. This approach identified ROIs in each subgroup that were either recruited for cognitive compensation (significantly increased BOLD activation after exercise) or had reduced activation indicating exercise-induced cognitive dysfunction and exertional exhaustion. Each ROI was tested in post hoc fashion in all subgroups for differences between pre-exercise and post-exercise BOLD response. The pre-exercise data determined if there were significant pre-existing differences in BOLD activation between groups. The post-exercise comparison determined if exercise caused significant differences in BOLD activation between subgroups.

Methods and materials

Ethics

The protocol was approved by the Georgetown University Institutional Review Board (IRB 2009-229, 2013-0943 and 2015-0579) and Department of Defense Congressionally Directed Medical Research Program (CDMRP) Human Research Program Office (HRPO) (A-15547 and A-18479), and listed in clinicaltrials.gov (NCT01291758 and NCT00810225). All clinical investigations were conducted according to the principles expressed in the Declaration of Helsinki.

Subjects

Veterans of the 1990–91 Persian Gulf War with GWI, and healthy, non-military control subjects were recruited to this 4-day long in-patient study in the Clinical Research Unit of the Georgetown–Howard Universities Center for Clinical and Translational Science. From the original study, we included the data of 10 control (5 deployed Gulf War-era veterans), 10 GWI-START, 18 GWI-STOOP and no POTS subjects (Rayhan et al., 2013). The additional recruits included 21 control (3 deployed veterans), 13 GWI-START, 28 GWI-STOOP and 11 GWI-POTS subjects (Garner et al., 2018). Subjects had history and physical examinations to ensure their inclusion by meeting Chronic Multisymptom Illness (Fukuda et al., 1998) and Kansas (Steele, 2000) criteria for GWI, confirmation of sedentary lifestyle for control subjects (<40 min of aerobic activity per week) and exclusion because of medical or psychiatric conditions (Steele, 2000; Reeves et al., 2003; Jones et al., 2009). Subjects completed Chronic Fatigue Syndrome Symptom Severity (Baraniuk et al., 2013), SF-36 quality of life (Ware and Sherbourne, 1992), Chalder Fatigue (Cella and Chalder, 2010) and McGill Pain (Dworkin et al., 2009) questionnaires and were examined for systemic hyperalgesia by dolorimetry (Naranch et al., 2002).

Protocols

Exercise

Two submaximal bicycle exercise tests were performed 24 h apart. Subjects cycled for 25 min at 70% predicted maximum HR (220-patient’s age), followed by a climb to 85% maximum HR to reach anaerobic threshold (Garner et al., 2018).

Postural tachycardia protocol and exercise-induced postural tachycardia subgroups

All subjects were assessed for postural tachycardia before and after exercise (Garner et al., 2018). Subjects rested supine for 5 min and had HR measured by continuous ECG monitor. After standing up, HR and blood pressure were measured every minute for 5 min. The incremental changes in HR after standing up (ΔHR) were calculated to identify episodes of postural tachycardia defined as ΔHR ≥30 bpm. The normal ΔHR was 12 ± 5 bpm (mean ± SD).

POTS was defined by ΔHR ≥30 bpm at two or more measurements prior to exercise during the 5-min standing periods. POTS criteria were met before and after exercise. Exercise did not exacerbate the postural tachycardia, so POTS subjects were distinct from the START and STOPP groups.

START was defined by having a normal ΔHR before exercise, but at least two measurements of postural tachycardia with ΔHR ≥30 bpm after exercise. This phenomenon was transient as ΔHR returned to normal 24–48 h following exercise (Garner et al., 2018).

STOPP was defined by normal ΔHR (ΔHR <30 bpm) at all times before and after exercise.
Verbal working memory task
Each subject performed the continuous 2-Back verbal working memory task during fMRI data acquisition (Fig. 1). Each 18-s stimulus block had instruction (REST, 0-BACK, 2-BACK, 2 s each), fixation (i.e. cross-hairs, 8 s), 0-Back and 2-Back components. For the 0-Back and 2-Back tasks, subjects viewed nine individual, pseudo-randomly selected upper-case letters (A, B, C, D) for 0.8 s each followed by an inter-stimulus interval (blank screen) of 1.2 s. For the 0-Back task, they pressed a button corresponding to the letter viewed. For the 2-Back task, they remembered the first two letters, then began to press the button corresponding to the letter seen two presentations (4 s) previous. These blocks were repeated for five cycles. Responses were collected from an MRI-compatible fibre-optic button box that subjects held with both hands. Response times between onset of each letter presentation and button pressing were recorded. There were 45 button responses for the 0-Back and 35 for the 2-Back tasks. The complete task (i.e. all five cycles) lasted a total of 5 min (300 s).

MRI data acquisition
All structural and functional MRI data were acquired on a Siemens 3T Tim Trio scanner located within the Center for Functional and Molecular Imaging at Georgetown University Medical Center equipped using a transmit-receive body coil and a commercial 12-element head coil array. Structural 3D T1-weighted Magnetization Prepared Rapid Acquisition Gradient Echo images parameters were: TR/TE = 1900/2.52 ms, flip angle = 9°, TI = 900 ms, FoV = 250 mm, 176 slices, slice resolution = 1.0 mm and voxel size = 1 × 1 × 1 mm. Magnetization Prepared Rapid Acquisition Gradient Echoes were all processed in SPM12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12/; Friston et al., 1996). fMRI data consisted of T2*-weighted gradient echo-planar images (EPIs) acquired during the 5-min task. EPI data acquisition parameters were: TR/TE = 2500/30 ms, flip angle = 90°, FoV = 205 mm², matrix size = 64 × 64, number of slices = 47 and voxel size = 3.2 mm³ (isotropic).

fMRI statistical analysis
Both the first-level (fixed effects) and second-level (random effects) statistical analyses were performed in SPM12.

First level
The first six scans from each block were removed to account for T1 saturation (Fig. 1). Data were pre-processed through the default pipeline within the CONN version 17 toolbox (https://web.conn-toolbox.org/; Whitefield-Gabrieli and Nieto-Castanon, 2012). In brief, steps were: (i) slice-timing correction, (ii) realignment and unwarping of the functional images, (iii) outlier detection for ‘scrubbing’ based on artefact detection tools, (iv) segmentation and spatial normalization of the anatomic scan to the Montreal Neurological Institute (MNI) template and (v) spatial smoothing with a stationary Gaussian filter of 6 mm full width at half maximum. Voxel size was 2.0 mm³ (isotropic) after spatial normalization. Pre-processed EPI data from each individual were modelled with the following events: instruction, fixation, 0-Back and 2-Back. The contrast of interest, 2-Back > 0-Back, was modelled using a one-sample t-test with motion parameters (i.e. translation and rotation) included as covariates of no interest. The resulting 2-Back > 0-Back contrast maps from every subject were then used in second-level analyses.

Second level
To identify regions where BOLD response changes related to working memory were significantly altered by exertion, for each subgroup we performed two-tailed, paired t-tests for 2-Back > 0-Back and 2-Back < 0-Back contrasts [cluster level: P < 0.05, family-wise error correction (FWE), kE ≥ 50]. Each ROI was then applied to all subgroups. The mean BOLD response within each significant ROI was extracted from the contrast maps of all subjects using MarsBaR 0.44 (https://www.fil.ion.ucl.ac.uk/spm/ext/#MarsBar; Brett et al., 2002). MarsBaR output was compared between the HC, GWI-START, GWI-STOPP and GWI-POTS groups on the pre- and post-exercise days by one-way ANOVA (P < 0.05) followed by post hoc Tukey’s Honestly Significant Difference (HSD) to correct for multiple comparisons. Exercise effects within each group were determined from 2-tailed paired t-tests between pre- and post-exercise data.

To minimize the effects of ‘double-dipping’ from the use of the same data set for ROI selection (Kriegeskorte et al., 2009; Button, 2019), significant HSDs between contrast groups that yielded ROIs were disregarded and FWE cluster-level significance was reported instead. Effect sizes for changes between the pre- and post-exertional fMRI sessions were estimated by Cohen’s d (paired studies) (Cohen, 1988, 1992; Ellis, 2010) and between groups on pre- and post-exercise days by Hedges’ g (Stangroom, 2019). In brief, effect sizes larger than 0.8 indicate strong studies and allow new studies of appropriate numbers of subjects to reproduce the effects in significant fashion (P < 0.05) with adequate power (>0.8) and so control false-positive and false-negative rates. These analyses were performed in SPSS (version 25).

The precise location and ‘regional compositional percentage’ (e.g. 10% anterior cingulate, 13% medial frontal gyrus, etc.) of each ROI in cortical, subcortical and cerebellar structures were found using a custom MATLAB script that employed functions from both SPM and xjView 9.6 (http://www.alivelearn.net/xjview/). Anatomical labels for every ROI voxel were identified using the AAL (Automated Anatomic Labelling) atlas. Cerebellar regions were confirmed with the spatially...
unbiased infra-tentorial template (SUIT) atlas (Diedrichsen, 2006; Diedrichsen et al., 2009).

Data availability
Data are available from the authors upon request and will be uploaded to OpenfMRI and/or the Boston Biorepository, Recruitment, and Integrative Network (BBRAIN) for GWI.

Results
Demographics
Age, percentage of males and body mass index were not different between groups (Table 1). GWI-START (n = 23), GWI-STOOP (n = 46) and GWI-POTS (n = 11) groups had equivalent results that, when combined, were significantly different from HC (n = 31) for CFS Symptom Severity Questionnaire, pain (McGill Total Pain Score) and tenderness (dolorimetry), fatigue (Chalder Fatigue Score) and quality of life (SF-36; ANOVA P < 0.004; HSD P < 0.05). All three GWI subgroups had significantly greater pain and tenderness than HC.

N-Back task performance
There were no significant differences between groups for response times or numbers of correct responses on the 0-Back or 2-Back tasks either before or after exercise (Fig. 2). Response times were faster for the 2-Back than 0-Back task.
In cortex, all groups activated the frontal-parietal executive control network before and after exercise (Fig. 3). Likewise, all groups activated several regions within the cerebellum before and after exercise, but a marked decline in cerebellar activity is evident after exercise in the GWI-START group (Fig. 4). As the GWI-POTS group had fewer subjects relative to the other groups, we kept height thresholds in Figs 3 and 4 at \( P < 0.001 \) (uncorrected) for visualization purposes.

ROI analysis identified three clusters that were significantly altered \( (P < 0.05, \text{FWE}) \) for visualization purposes. The larger of the two cerebellar ROIs (186 voxels) was in the right hemisphere and included voxels in the right cerebellar lobule VI (43/186 voxels, 23%), dentate nucleus (43/186, 23%), lobule IV (7/186, 4%) and lobule III (5/186, 3%; Fig. 5A). The remaining voxels (88/186) were undefined in the AAL and SUIT atlases. Activation in this ROI was equivalent between groups prior to exercise (Fig. 5B). This ROI resulted from deactivation after exercise in the GWI-START group (cluster level: \( P = 0.0028, \text{FWE} \); Fig 5C). Post hoc analyses examined differences between subgroups on each day. Prior to exercise, the HC, GWI-START, GWI-STOPP and GWI-POTS groups had equivalent activation levels \( \text{ANOVA: } F(3, 110) = 0.343, P = 0.80 \). After exercise, levels were significantly different \( \text{ANOVA: } F(3, 110) = 5.64, P = 0.001 \), and GWI-START had significant deactivation relative to HC (HSD, \( P = 0.001 \)) and GWI-STOPP (HSD, \( P = 0.041 \)). Paired comparisons of pre- versus post-exertional activity revealed a trend towards deactivation in GWI-POTS (paired \( t \)-test, \( P = 0.051 \); Fig. 5D). The effects of exercise were large with Cohen’s \( d \) of 1.7 for GWI-START and 0.89 for GWI-POTS. Effect sizes were small for GWI-STOPP \( (d = 0.28) \) and HC \( (d = 0.02) \). Hedges’ \( g \) indicated strong effect sizes to detect differences between GWI-START and the post-exercise HC \( (g = 1.16) \), GWI-STOPP \( (g = 0.92) \) and GWI-POTS \( (g = 0.73) \) results (Stangroom, 2019).

The second cerebellar ROI (145 voxels) was more medioposterior and included vermis lobules VII (51/145 voxels, 35%) and VI (30/145, 21%), and right cerebellar lobule VI (27/145, 19%; Fig. 6A). Activation in this second cerebellar ROI closely paralleled that of the first, with equivalent activity between groups prior to exercise (Fig. 6B) and deactivation after exercise in the GWI-START group (cluster level: \( P = 0.011, \text{FWE} \)) with a similar trend in GWI-POTS \( (P = 0.054 \); Fig. 6C). The responses before exercise were similar \( \text{ANOVA: } F(3, 110) = 0.648, P = 0.65 \). The responses after exercise

**Table 1** Demographics

|                  | HC     | GWI-START | GWI-STOPP | GWI-POTS |
|------------------|--------|-----------|-----------|----------|
| \( N \)          | 31     | 23        | 46        | 11       |
| Age (yr)         | 43.9 ± 15.9 | 45.1 ± 8.1 | 48.6 ± 7.4 | 45.7 ± 6.0 |
| Male             | 0.645 ± 0.486 | 0.818 ± 0.395 | 0.756 ± 0.435 | 0.818 ± 0.405 |
| Body mass index  | 28.7 ± 4.5 | 27.9 ± 4.3 | 29.8 ± 5.9 | 30.4 ± 6.1 |
| McGill total pain score | 4.2 ± 7.0  | 26.8 ± 8.6 | 23.2 ± 8.7 | 23.0 ± 10.8 |
| Dolorimetry (kg) | 6.0 ± 1.9*  | 3.3 ± 2.1  | 3.6 ± 1.9  | 4.0 ± 2.3  |
| Chalder Fatigue Score | 12.4 ± 5.4* | 26.1 ± 4.4 | 25.6 ± 4.7 | 24.8 ± 5.3 |
| CFSS Questionnaire |        |           |           |          |
| Fatigue          | 1.4 ± 1.3*  | 3.8 ± 0.4  | 3.4 ± 0.8  | 3.4 ± 0.7  |
| Short term memory | 1.3 ± 1.3*  | 3.2 ± 0.9  | 3.2 ± 0.8  | 2.9 ± 0.7  |
| Sore throat      | 0.3 ± 0.7*  | 1.9 ± 1.2  | 1.3 ± 1.2  | 1.3 ± 1.2  |
| Sore lymph nodes | 0.1 ± 0.4*  | 2.0 ± 1.3  | 1.3 ± 1.2  | 1.6 ± 1.2  |
| Muscle pain      | 0.8 ± 1.1*  | 3.4 ± 0.6  | 3.1 ± 1.0  | 3.0 ± 1.2  |
| Joint pain       | 0.9 ± 1.2*  | 3.5 ± 0.9  | 3.2 ± 0.9  | 2.8 ± 1.1  |
| Headaches        | 1.0 ± 1.2*  | 2.9 ± 1.1  | 2.6 ± 1.2  | 3.0 ± 1.1  |
| Sleep            | 1.8 ± 1.4*  | 3.7 ± 0.5  | 3.5 ± 0.8  | 3.2 ± 1.0  |
| Exertional exhaustion | 0.7 ± 1.2* | 3.8 ± 0.5  | 3.2 ± 1.0  | 3.4 ± 1.0  |
| SF-36 Physical functioning | 83.8 ± 24.4* | 42.4 ± 25.7 | 47.0 ± 22.6 | 45.5 ± 22.7 |
| Role physical    | 76.7 ± 39.4* | 60.0 ± 22.2 | 11.3 ± 25.3 | 0.0 ± 0.0  |
| Bodily pain      | 80.9 ± 21.2* | 22.8 ± 16.5 | 28.9 ± 18.4 | 34.1 ± 15.0 |
| General health   | 67.6 ± 23.3* | 20.7 ± 15.5 | 29.0 ± 20.3 | 23.4 ± 16.4 |
| Vitality         | 58.0 ± 23.0* | 13.3 ± 11.0 | 17.9 ± 17.1 | 20.5 ± 13.9 |
| Social functioning | 77.1 ± 27.7* | 22.6 ± 17.9 | 33.1 ± 25.4 | 33.0 ± 18.8 |
| Role emotional   | 84.4 ± 32.4* | 25.4 ± 36.4 | 30.8 ± 39.5 | 30.3 ± 40.7 |
| Mental health    | 73.9 ± 16.7* | 49.7 ± 21.5 | 55.5 ± 22.3 | 54.5 ± 23.7 |

*ANOVA with \( P < 0.05 \) followed by \( P < 0.05 \) by HSD for HC compared with the combined GWI subgroups.

\[ P = 0.0028, \text{FWE} \; \text{Fig} \; 5C \]. Post hoc analyses examined differences between subgroups on each day. Prior to exercise, the HC, GWI-START, GWI-STOPP and GWI-POTS groups had equivalent activation levels \( \text{ANOVA: } F(3, 110) = 0.343, P = 0.80 \). After exercise, levels were significantly different \( \text{ANOVA: } F(3, 110) = 5.64, P = 0.001 \), and GWI-START had significant deactivation relative to HC (HSD, \( P = 0.001 \)) and GWI-STOPP (HSD, \( P = 0.041 \)). Paired comparisons of pre- versus post-exertional activity revealed a trend towards deactivation in GWI-POTS (paired \( t \)-test, \( P = 0.051 \); Fig. 5D). The effects of exercise were large with Cohen’s \( d \) of 1.7 for GWI-START and 0.89 for GWI-POTS. Effect sizes were small for GWI-STOPP \( (d = 0.28) \) and HC \( (d = 0.02) \). Hedges’ \( g \) indicated strong effect sizes to detect differences between GWI-START and the post-exercise HC \( (g = 1.16) \), GWI-STOPP \( (g = 0.92) \) and GWI-POTS \( (g = 0.73) \) results (Stangroom, 2019).
were significantly different from each other [ANOVA: $F(3, 110) = 5.21, P = 0.0087$], with negative BOLD signal in GWI-START compared with GWI-STOOP (HSD, $P = 0.008$). Thus, these results like those for the larger cerebellar ROI reflect a significant relative decrease in GWI-START activity, not an increase in HC and GWI-STOOP activity, after exercise.

One GWI-START subject was an outlier for the cerebellar ROIs following exercise. Pre-exercise values were comparable with all other subjects. However, after exercise, the BOLD response increased to more than two standard deviations greater than the mean for the other GWI-START group members, reaching 1.26 and 0.61 for the 145- and 186-voxel ROIs, respectively. This was a focal finding because all of this subject’s other BOLD responses were comparable with the other GWI-START subjects. No abnormality was seen on structural scans. The subject’s data were excluded from the analysis of the cerebellar clusters but were used for all other analyses.

The third ROI (Fig. 7) was characterized by a significant increase in activation in the GWI-STOOP group following exercise (cluster-level $P = 0.0068$, FWE; Cohen’s $d = 0.55$). The 191-voxel ROI mapped to the anterior right supplementary motor area (163/191, 85%). MNI co-ordinates for this third ROI overlap with the dorsal anterior cingulate cortex regions of the anterior salience network as described by Shirer et al. (2012). Activation in this ROI was equivalent between groups before and after exercise. HC, GWI-START, GWI-STOOP and GWI-POTS groups had similar activation levels before [ANOVA: $F(3, 110) = 2.08, P = 0.11$] and after [ANOVA: $F(3, 110) = 1.47, P = 0.23$] exercise. GWI-STOOP was predicted to have higher BOLD responses following exercise provocations in future studies when compared with GWI-START (Hedges’ $g = 1.01$) and GWI-POTS ($g = 0.73$).

Table 2 contains regional comparisons between the AAL and SUIT atlases.

**Discussion**

Exercise caused alterations of regional brain blood flow measured by BOLD activity during the continuous
N-back working memory task (2-Back > 0-Back condition), and autonomic changes leading to the transient postural tachycardia of the START group. Before exercise, BOLD responses were positive and equivalent among the four groups. Exercise had no effect on BOLD responses in HC subjects.

GWI-START had significant deactivation of right cerebellar dentate nucleus and vermis lobules VI and VII (Figs 5 and 6). GWI-POTS had a trend for deactivation in the cerebellar ROIs after exercise ($P < 0.054$ by paired $t$-tests, Fig. 5) that was limited by the small sample size ($n = 11$, Cohen’s $d = 0.89$). The cerebellar findings (Figs 5 and 6) confirmed previous studies showing activation of cerebellar lobules VI, VII (Stoodley et al., 2012), crus I and dentate nucleus (Thurling et al., 2012) with various other N-Back task designs. Increasing task difficulty and working memory cognitive load recruit additional areas in the dentate nucleus and vermis VI and VII for cognitive compensation (Kuper et al., 2016). Roles for cerebellar dentate and vermis lobules VI and VII in verbal working memory are well attested from numerous brain lesion and neuroimaging studies (Desmond et al., 1997; Chen and Desmond, 2005; Kirschen et al., 2005; Hautzel et al., 2009; Stoodley and Schmahmann, 2009; Kirschen et al., 2010; Cooper et al., 2012; Stoodley et al., 2012; Thurling et al., 2012). Other roles include sensorimotor control (Sokolov et al., 2017), reward circuits, social behaviour and dysfunction in autism spectrum disorder, schizophrenia and addiction (Carta et al., 2019).

GWI-STopp had a gain of function response after exercise with a significant increase in BOLD response in the right anterior supplementary motor area ($F = 0.053$ by paired $t$-tests, Fig. 7). The MNI co-ordinates for this right anterior supplementary motor area ROI overlap with one atlas’s dorsal anterior cingulate cortex subdivision of the anterior salience.
network (Shirer et al., 2012). This ROI is also immediately adjacent to ‘pre-supplementary motor area’ and ‘midcingulate’ regions that are activated during pain in HC children (Hohmeister et al., 2010) and by thermal pain during a motor task in young adults (Misra and Coombes, 2015). The latter study found functional connectivity between the supplementary motor area and cerebellar lobules VI and VII indicating simultaneous multimodal processing of motor control and pain (Misra and Coombes, 2015; Coombes and Misra, 2016).

Figure 5 Right cerebellar cluster. The 186-voxel ROI was identified from the exercise-induced decrease in BOLD activation in the GWI-START subgroup (cluster-level \( p = 0.0028 \), FWE). (A) Sagittal (top), coronal (middle) and transverse (bottom) slices of an MNI-standard brain, where cross-hairs indicate the cluster’s most active voxel (24, –46, –32). BOLD response for the 2-Back > 0-Back condition (mean ± SEM) are shown for (B) pre-exercise and (C) post-exercise changes in BOLD responses for the control (white bars), GWI-START (black bars), GWI-STopp (light grey bars) and GWI-POTS (dark grey bars). The ANOVA \( P \)-values refer to the difference between the HC and the combined three GWI subgroups. (D) Post-minus pre-exercise BOLD response for the 2-Back > 0-Back condition for the HC, GWI-START, GWI-STopp and GWI-POTS groups.
Limitations

The quantitative ROI analysis verified deactivation in the cerebellar vermis in the GWI-START group after exertion but did not substantiate the qualitative impression of basal ganglia and anterior insula activation in the GWI-STOPP group that was suggested by the earlier interim analysis (Rayhan et al., 2013). The exercise was required to unveil the transient postural tachycardia in GWI-START and patterns of BOLD activation in GWI subsets, indicating that resting state or baseline studies may not distinguish between GWI-START, GWI-STOPP and control groups. GWI-POTS subjects had cerebellar

Figure 6 Medial cerebellar cluster. The 145-voxel ROI was identified from the exercise-induced incremental BOLD deactivation in the GWI-START subgroup (cluster-level $P = 0.011$, FWE). (A) Sagittal (top), coronal (middle) and transverse (bottom) slices of an MNI-standard brain, where cross-hairs indicate the cluster’s most active voxel ($10, -70, -26$). BOLD response for the 2-Back > 0-Back condition (mean ± SEM) are shown for (B) pre-exercise and (C) post-exercise changes in BOLD responses for the control (white bars), GWI-STOPP (light grey bars), GWI-POTS (dark grey bars) and GWI-START (black bars). The ANOVA $P$-values refer to the difference between the HC and the combined three GWI subgroups. (D) Post-minus pre-exercise BOLD response for the 2-Back > 0-Back condition for the HC, GWI-START, GWI-STOPP and GWI-POTS groups.
deactivation after exercise comparable to the GWI-START group, but larger studies will be needed to examine the statistical significance. Effects sizes were provided to help power future studies.

The study spanned two campaigns of recruitment and testing, but the exercise, N-Back and analytical procedures were identical. All scans were analysed as a single batch with extensive computational efforts to reduce the effects of head motion. This was a particular problem for veterans with severe pain and substantial systemic hyperalgesia.

Our ‘continuous’ version of the N-Back task had the unusual effect of eliciting faster response times from

Figure 7. Right Supplementary Motor Area Cluster. The 191-voxel ROI was identified from the exercise-induced increased BOLD activation in the GWI-STOpP subgroup (cluster-level: \( P = 0.0068 \), FWE). (A) Sagittal (top), coronal (middle) and transverse (bottom) slices of an MNI-standard brain, where cross-hairs indicate the cluster’s most active voxel (8, 14, 54). BOLD response for the 2-Back > 0-Back condition (mean ± SEM) are shown for (B) pre-exercise and (C) post-exercise changes in BOLD responses for the control (white bars), GWI-START (black bars), GWI-POTS (dark grey bars) and GWI-STOpP (light grey bars). The ANOVA \( P \)-values refer to the difference between the HC and the combined three GWI subgroups. (D) Post-minus pre-exercise BOLD response for the 2-Back > 0-Back condition for the HC, GWI-START, GWI-STOpP and GWI-POTS groups.
Table 2 Comparison of cerebellar ROI voxel co-ordinate labels in the SUIT and AAL atlases

| Cerebellar ROI (Peak Voxel) | SUIT labels | AAL labels | Voxel count | Voxel count |
|-----------------------------|-------------|------------|-------------|-------------|
| 145-Voxel Cluster (10, -70, -26) | Left | Left | 3 | 4 |
| | Crus I | Lobule 8 | 6 | 8 |
| | Lobule VI | Crus I | 4 | 3 |
| | Vermis Crus II | Crus2 | 6 | 8 |
| | Vermis VI | Vermis 7 | 6 | 9 |
| | Vermis VIIb | Vermis 7 | 4 | 9 |
| | Crus I | Lobule VI | 11 | 27 |
| | Crus II | Lobule 8 | 1 | 1 |
| | Lobule VI | Crus I | 36 | 13 |
| | Vermis Crus II | Crus2 | 7 | 4 |
| | Vermis VI | Vermis 6 | 64 | 30 |
| | Right | Vermis 7 | 42 | 42 |
| | Undefined | Left | 2 | 4 |
| | Lobule V | Vermis 4/5 | 1 | 1 |
| | Right | Right | Left | 1 | 1 |
| | Dentate | Lobule 3 | 7 | 5 |
| | Lobule I/IV | Lobule 4/5 | 34 | 43 |
| | Lobule V | Vermis 1/2 | 21 | 4 |
| | Lobule VI | Vermis 4/5 | 26 | 26 |
| | Right | Vermis 6 | 1 | 1 |
| | Undefined | Undefined | 97 | 96 |
| 186-Voxel Cluster (24, -46, -32) | Left | Left | 3 | 4 |
| | Vermis Crus II | Crus2 | 6 | 8 |
| | Vermis VI | Vermis 7 | 6 | 9 |
| | Vermis VIIb | Vermis 7 | 4 | 9 |
| | Crus I | Lobule VI | 11 | 27 |
| | Crus II | Lobule 8 | 1 | 1 |
| | Lobule VI | Crus I | 36 | 13 |
| | Vermis Crus II | Crus2 | 7 | 4 |
| | Vermis VI | Vermis 6 | 64 | 30 |
| | Right | Vermis 7 | 42 | 42 |
| | Defined | Left | 2 | 4 |
| | Lobule V | Vermis 4/5 | 1 | 1 |
| | Right | Right | Left | 1 | 1 |
| | Dentate | Lobule 3 | 7 | 5 |
| | Lobule I/IV | Lobule 4/5 | 34 | 43 |
| | Lobule V | Vermis 1/2 | 21 | 4 |
| | Lobule VI | Vermis 4/5 | 26 | 26 |
| | Right | Vermis 6 | 1 | 1 |
| | Defined | Undefined | 97 | 96 |

Two groups, and to determine if there were significant intergroup differences on the pre- and post-exercise days. Furthermore, our ROI analyses employed only contiguous clusters, which are less susceptible to serious result distortion from selective analyses of selected data (Kriegeskorte et al., 2009). Critics of circular analysis in systems neuroscience concede that selective in-depth analysis of ROIs can provide additional insights to analyses employing non-selective mapping (Saxe et al., 2006; Kriegeskorte et al., 2009). Lastly, as we have addressed the need for larger studies, one can leverage larger subject pools to address issues with double-dipping. Specifically, provided a sufficiently large population, one can divide the population into two contrast groups, one for ROI detection and the other for quantitation within that ROI (Kriegeskorte et al., 2009).

The entire period of the two submaximal bicycle exercise stress test, postural tachycardia testing and overnight stay in the Clinical Research Unit constituted the physiological stressor. Our design did not determine if a single exercise stress test was sufficient to alter BOLD responses. Regional cerebellar deactivation was inferred to indicate cognitive decompensation, while the enhanced activation in the right anterior supplementary motor area may indicate cognitive compensation by the GWI-STOPP group.

Lastly, the data presented here are restricted to the BOLD responses elicited by verbal working memory (i.e. 2-Back > 0-Back). It is possible to differentiate veterans with GWI from HCs using other neural processes. For example, contrasting 0-Back > 2-Back revealed that the magnitude of deactivation in the default mode networks of veterans with GWI is greater than that of HCs (Rayhan et al., 2019). Thus, we plan future studies that will report the results of other contrasts in these data (e.g. 0-Back > 2-Back, 2-Back > Fixation, etc.) so that we can better differentiate autonomic subgroups of GWI from each other and HCs.

Conclusion

The exercise stressor protocol caused distinct changes in the patterns of dysfunction in veterans with GWI. After exercise, GWI-STOPP activated the right supplementary motor area/dorsal anterior cingulate cortex while GWI-START had cerebellar deactivation. GWI-POTS had a trend towards cerebellar deactivation similar to GWI-START. Thus, patterns of cerebellar and cortical somatosensory activation could be used to quantitatively distinguish the GWI-START from the GWI-STOPP phenotype. These results suggest that the exercise stressor paradigm affected different neural mechanisms in the GWI phenotypes and that distinct neural networks may mediate the cognitive dysfunction, post-exertional malaise and exertional exhaustion in these veterans.
Acknowledgements

We would like to acknowledge our participants and their families, whose co-operation made this research possible. We would also like to acknowledge Drs. Adeen Flinker and Kyle Shattuck for helpful advice regarding data analyses.

Funding

The study was supported by funding from the Sergeant Sullivan Circle, Dr. Barbara Cottone, Dean Clarke Bridge Prize, Department of Defense Congressional Directed Medical Research Program (CDMRP) W81XWH-15-1-0679 and W81-XWH-09-1-0526, and the National Institute of Neurological Disorders and Stroke R21NS088138 and RO1NS085131. This project has been funded in whole or in part with Federal funds (Grant #UL1TR000101 previously UL1RR031975) from the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH), through the Clinical and Translational Science Awards Program (CTSA), a trademark of DHHS, part of the Roadmap Initiative, ‘Re-Engineering the Clinical Research Enterprise’.

Competing interests

The authors report no competing interests.

References

Baraniuk JN, Adewuyi O, Merck SJ, Ali M, Ravindran MK, Timbol CR, et al. A Chronic Fatigue Syndrome (CFS) severity score based on case designation criteria. Am J Transl Res 2013; 5: 53–68.

Brett M, Anton J-L, Valabregue R, Poline J. Region of interest analysis using an SPM toolbox. In: 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan, 2002.

Butler KS. Double-dipping revisited. Nat Neurosci 2019; 22: 688–90.

Carta I, Chen CH, Schott AL, Dorizan S, Khodakhah K. Cerebellar modulation of the reward circuitry and social behavior. Science 2019; 363: eaav0581.

Cella M, Chalder T. Measuring fatigue in clinical and community settings. J Psychosom Res 2010; 69: 17–22.

Chen SH, Desmond JE. Cerebrocerebellar networks during articulatory rehearsal and verbal working memory tasks. Neuroimage 2003; 24: 332–8.

Cohen J. Statistical power analysis for the behavioral sciences. 2nd edn. Hillsdale, NJ: L. Erlbaum Associates; 1988.

Cohen J. A power primer. Psychol Bull 1992; 112: 155–9.

Demirakca S, Flor H, et al. Cerebral processing of pain in school-aged children with neonatal nociceptive input: an exploratory fMRI study. Pain 2016; 157: 117–27.

De Ridder D, Vanneste S, Freeman W. The Bayesian brain: phantom percepts resolve sensory uncertainty. Neurosci Biobehav Rev 2014; 44: 4–15.

Desmond JE, Gabrieli JD, Wagner AD, Ginier BL, Glover GH. Lobular patterns of cerebellar activation in verbal working-memory and finger-tapping tasks as revealed by functional MRL. J Neurosci 1997; 17: 9675–85.

Diedrichsen J. A spatially unbiased atlas template of the human cerebellum. Neuroimage 2006; 33: 127–38.

Diedrichsen J, Balsters JH, Flavel J, Cussans E, Ramnani N. A probabilistic MR atlas of the human cerebellum. Neuroimage 2009; 46: 59–46.

Dursa EK, Barth SK, Schneiderman AI, Bossarte RM. Physical and mental health status of Gulf War and Gulf Era veterans: results from a large population-based epidemiological study. J Occup Environ Med 2016; 58: 41–6.

Dworkin RH, Turk DC, Revicki DA, Harding G, Coynes KS, Peirce-Sandner S, et al. Development and initial validation of an expanded and revised version of the Short-form McGill Pain Questionnaire (SF-MPQ-2). Pain 2009; 144: 35–42.

Ellis PD. The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results. Cambridge and New York: Cambridge University Press; 2010.

Fiedler N, Ozakinci G, Hallman W, Wartenberg D, Brewer NT, Barrett DH, et al. Military deployment to the Gulf War as a risk factor for psychiatric illness among US troops. Br J Psychiatry 2006; 188: 453–9.

Friston KJ, Williams S, Howard R, Frackowiak RS, Turner R. Movement-related effects in fMRI time-series. Magn Reson Med 1996; 35: 346–55.

Fukuda K, Nisenbaum R, Stewart G, Thompson WW, Robin L, Washiko RM, et al. Chronic multisymptom illness affecting Air Force veterans of the Gulf War. JAMA 1998; 280: 981–8.

Garner RS, Rayhan RU, Baraniuk JN. Verification of exercise-induced transient postural tachycardia phenotype in Gulf War Illness. Am J Transl Res 2018; 10: 3254–64.

Georgopoulos AP, James LM, Carpenter AF, Engdahl BE, Leuthold AG, Lewis SM. Gulf War Illness (GWI) as a neuroimmune disease. Exp Brain Res 2017; 235: 3217–25.

Hautzel H, Mottaghy FM, Specht K, Muller HW, Krause BJ. Evidence of a modality-dependent role of the cerebellum in working memory? An fMRI study comparing verbal and abstract n-back tasks. Neuroimage 2009; 47: 2073–82.

Hohmeister J, Kroll A, Wollgarten-Hadamek I, Zohsel K, Demirakca S, Flor H, et al. Cerebral processing of pain in school-aged children with neonatal nociceptive input: an exploratory fMRI study. Pain 2010; 150: 257–67.

Jamulewicz PA, Krengel MH, Maule A, White RF, Cirillo J, Sisson E, et al. Neuropsychological characteristics of Gulf War illness: a meta-analysis. PLoS One 2017; 12: e0177121.

Jones JF, Lin JM, Maloney EM, Boneva RS, Nater UM, Unger ER, et al. An evaluation of exclusionary medical/psychiatric conditions in the definition of chronic fatigue syndrome. BMC Med 2009; 7: 57.

Kirsch MN, Chen SH, Desmond JE. Modality specific cerebro-cerebellar activations in verbal working memory: an fMRI study. Behav Neurology 2010; 23: 51–63.

Kirsch MN, Chen SH, Schraedley-Desmond P, Desmond JE. Load- and practice-dependent increases in cerebro-cerebellar activation in verbal working memory: an fMRI study. Neuroimage 2005; 24: 462–72.

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 2009; 12: 340–50.

Kuper M, Kaschani P, Thurling M, Stefanescu MR, Burciu RG, Goricke S, et al. Cerebellar fMRI activation increases with increasing working memory demands. Cerebellum 2016; 15: 322–35.

Misra G, Coombes SA. Neuroimaging evidence of motor control and pain processing in the human midcingulate cortex. Cereb Cortex 2015; 25: 1906–19.

Naranch K, Park YJ, Repka-Ramirez MS, Velarde A, Clauw D, Baraniuk JN. A tender sinus does not always mean rhinosinusitis. Otolaryngol Head Neck Surg 2002; 127: 387–97.
