Journal of the International Neuropsychological Society (1999), **5**, 203–212.

Neurobehavioral deficits in Persian Gulf veterans: Evidence from a population-based study

W. KENT ANGER,[1] DANIEL STORZBACH,[1,2] LAURENCE M. BINDER,[2] KEITH A. CAMPBELL,[2] DIANE S. ROHLMAN,[1] LINDA MCCAULEY,[1] CRAIG A. KOVERA,[1] KELLY L. DAVIS,[1] AND THE PORTLAND ENVIRONMENTAL HAZARDS RESEARCH CENTER

[1]Occupational and Environmental Toxicology; Oregon Health Sciences University, Portland, OR 97201
[2]Portland VA Medical Center, Portland, OR 97207

(RECEIVED January 7, 1998; REVISED June 1, 1998; ACCEPTED June 15, 1998)

Abstract

Reports of low-concentration nerve gas exposures during the Persian Gulf War have spurred concern about possible health consequences and refocused interest on the symptoms reported by many returning military veterans. The Portland Environmental Hazards Research Center is studying veterans from the Northwest USA who report persistent, unexplained "Persian Gulf" symptoms (*cases*) or who do not report those symptoms (*controls*). Of the first 101 veterans studied, cases differed substantially from controls on a broad range of psychological tests indicative of increased distress. A subgroup of cases was identified with objective deficits on neurobehavioral tests of memory, attention, and response speed. (*JINS*, 1999, *5*, 203–212.)

Keywords: Neurobehavioral assessment, Persian Gulf War, PGW

INTRODUCTION

U.S. veterans who served in the 1991 Persian Gulf (PG) conflict have reported a wide variety of symptoms arising during their service in Southwest Asia or shortly after they returned. Symptoms associated with PG service ("PG symptoms") include memory and attention losses, fatigue, skin rash, muscle and joint pain, and gastrointestinal distress (Centers for Disease Control, 1995; Hyams et al., 1996; Iowa Persian Gulf Study Group, 1997). The U.S. Departments of Veterans Affairs (DVA) and Defense (DOD) have conducted extensive examinations of veterans reporting symptoms or other medical concerns. Those examinations have failed to yield a medical diagnosis or explanation for many veterans who report PG symptoms (Committee to Review the Health Consequences of Service During the Persian Gulf War; CRHCSPGW, 1996; Newmark & Clayton, 1995). Unfortunately, data generated from these evaluations have limited research value because the population is self-selected (Goldstein et al., 1996; Persian Gulf Veterans Coordinating Board; PGVCB, 1995).

One consistent opinion to emerge from research on PG veterans is that unexplained PG symptoms may reflect a variety of illnesses or disorders rather than a single, unique syndrome (Goldstein et al., 1996). Speculation that psychological factors and low-dose (especially organophosphate) chemical-warfare agent exposures could explain some of the PG symptoms (CRHCSPGW, 1996; PGVCB, 1995) has led to research assessing these factors. Axelrod and Milner (1996) reported Stroop and Grooved Pegboard performance deficits in 44 PG veterans from an Army National Guard unit, compared to population norms. Furthermore, the veterans with lower scores differed on psychological tests in the direction of increased distress. Sillanpaa et al. (1997) administered neuropsychological examinations to veterans participating in weekend clinics, most reporting PG symptoms. Regression analyses associated the symptoms with psychological and neuropsychological measures, leading the authors to conclude that the veterans' subjective complaints were related to emotional function. Goldstein et al. (1996) reported that 21 PG veterans who self-referred to the local PG registry (a DVA mechanism for veterans concerned about the development or possible development of health problems associated with their service in the PG) with complaints of PG symptoms did not differ from 21 matched symptom-free controls from the local community on indi-

Reprint requests to: W. Kent Anger, OHSU CROET L606, 3181 SW Sam Jackson Park Road, Portland, OR 97201. E-mail: anger@ohsu.edu

vidual neuropsychological tests in the Pittsburgh Occupational and Environmental Tests (POET) battery, (Ryan et al., 1987). However, a global index based on the number with tests 1 standard deviation below the normal mean did reveal a significant impairment in the PG group.

Haley et al. (1997), Haley and Kurt (1997) and Hom et al. (1997) administered an extensive battery of neuropsychological tests to 26 PG veterans with symptom complaints, compared to results of 20 veterans deployed to the PG and 10 not deployed. They reported that the group with symptoms had neuropsychological impairment that was in some cases indicative of brain damage and that the group "had lower abilities across a broad range of neuropsychological function including intelligence, academic abilities, higher cognitive function, language and visual–spatial function and sensorimotor abilities" (Hom et al., 1997, p. 540). Based on the sensory deficits and the similarity to test results in organophosphate-exposed poisoning victims, they hypothesize that "wartime exposure to neurotoxic chemicals is the causative factor" (Hom et al., 1997, p. 541), although no evidence of specific organophosphate exposure is identified for the 26 highly selected veterans who serve as the basis for their report.

Each of these reports is a clinical study of self-selected patients who reported PG symptoms. They are susceptible to the criticism that the sickest individuals were studied and not surprisingly differed from norms or community controls. It has been argued that such studies limit conclusions to the sample studied, not the population from which they emerged (e.g., Dikmen et al., 1992). The Portland (Oregon) Environmental Hazards Research Center (PEHRC) initiated in 1995 a population-based study of unexplained, nationally reported PG symptoms and factors that may contribute to their etiology. This report describes interim findings (testing completed as of December, 1996) from the PEHRC study component which seeks (1) to determine if symptomatic veterans have objective deficits in memory and attention and (2) to assess psychological conditions that might contribute to the presentation of the broad range of unexplained symptoms.

METHODS

Participant Pool

The U.S. Department of Defense (DOD) provided PEHRC with a list of all U.S. veterans who listed Oregon or Washington as their home state at the time of deployment to the Persian Gulf between September 1, 1990 and August 31, 1991. Beginning in December, 1995, questionnaires were mailed to 620 randomly selected members of this Northwest U.S. veteran population, and 385 were completed, a 62% response rate. The initial goal of the questionnaires was to identify asymptomatic veterans (potential controls) and those reporting at least one of the following symptoms (potential cases): (1) cognitive or psychological changes including memory loss, confusion, inability to concentrate, mood swings and/or somnolence; (2) gastrointestinal distress; (3) fatigue; (4) muscle and joint pain; and (5) skin or mucous membrane lesions. To qualify as a potential case, respondents had to affirm that the symptoms: (1) were persistent (lasting 1 month or longer); (2) began during or after service in the Gulf; and, (3) were present during the past 3 months (before the questionnaire arrived).

Questionnaire respondents were deemed ineligible for the study if they had served in the Vietnam conflict or if they refused further contact by study personnel. Veterans endorsing any of the PG symptoms (by answering *yes* to at least one of the multiple questions associated with each symptom) were contacted by telephone to verify the symptoms reported on the questionnaire and to search through self-reported medical history for a diagnosis that could credibly explain each symptom reported (e.g., head trauma, diabetes). Veterans reporting such diagnosed conditions were excluded *in order to focus the study on veterans with unexplained symptoms*.

Examinations

If no medical explanation for the symptoms was identified by telephone, potential cases and controls were recruited for testing. The 6-to-8-hr study protocol included a physical examination with emphasis on neurological and musculoskeletal systems, detailed health histories, blood and urine samples, and neurobehavioral and psychological tests. Respondents who did not endorse any PG symptom were designated potential controls and given the same neurobehavioral and psychological tests, provided blood and urine samples and were also screened by physicians for unrecognized signs or symptoms of illness. The physicians were blind to the potential case/control designation when administering the screening exams, as were the examiners when overseeing the neurobehavioral and psychological tests. Veterans completing the medical examination and neurobehavioral and psychological tests received $50 to defray expenses. At each stage of the investigation, informed consent was obtained after the nature and possible consequences of participation were explained.

Caseness Determination

Following medical examinations and any necessary referrals to identify a diagnostic explanation for the PG symptoms reported by the veteran, a clinical caseness determination committee reviewed each potential case. Committee membership included physicians and research scientists with doctoral degrees in neurology, rheumatology, internal medicine, neuropsychology, and epidemiology. Potential cases with explainable diagnoses or who denied PG symptoms at the time of clinical examination were excluded from the study *in order to focus on unexplained symptoms*. The primary bases for exclusion were the occurrence of diabetes mellitus or conditions uniquely associated with development of the five

main PG symptom categories under study by PEHRC (e.g., head injury for cognitive or psychological symptoms; scleroderma with accompanying skin rash for skin symptoms; gastric or duodenal ulcer for gastrointestinal symptoms; current pain in an area of previous surgery for muscle/joint symptoms; or abnormal thyroid hormone levels for fatigue symptoms). Clinical test results of potential controls were reviewed using the same exclusionary criteria applied to potential cases.

Psychological Tests

A 4-hr battery of 18 tests was used to assess psychological and neurobehavioral function, consisting of questionnaire measures of symptoms of Posttraumatic Stress Disorder (PTSD); health symptoms; personality; symptoms of psychopathology; combat and life experiences that might underlie other conditions; and objective measures of memory or concentration, complex cognitive processing, and response speed.

The psychological tests (questionnaires) were presented in the computerized Health Screening System (HSS; Kovera et al., 1996). The neurobehavioral tests were drawn from the Behavioral Assessment and Research System (BARS), a computer-implemented battery (Anger et al., 1994, 1996; Rohlman et al., 1996) of measures sensitive to neurotoxic insult (Anger, 1990). BARS and HSS tests were programmed in Allegiant SuperCard and C++, and presented on an Apple Powerbook 540c or 5300c with active matrix screens. Participants responsed on a durable nine-button (each measuring 20 × 15 mm) response unit termed the DataSled placed over the Powerbook keyboard (pictured in Kovera et al., 1996).

The 12 psychological tests implemented in the HSS for this study were as follows: (1) Mississippi PTSD Scale, a 39-item scale that reported war-related military posttraumatic stress disorder (PTSD) symptoms (Keane et al., 1988); (2) Penn Inventory for PTSD, a 26-item scale that tapped DSM–III–R PTSD symptoms (Hammarberg, 1992); (3) Posttraumatic Stress Disorder Checklist (PCL–C), a 17-item questionnaire that assessed frequency of complaints in relation to stressful experiences (Weathers et al., 1993); (4) Health Status Questionnaire (SF-36), a 36-item questionnaire that assessed functional (somatic) impairment and symptoms due to medical health problems (Ware & Sherbourne, 1992; Ware et al., 1988); (5) Beck Anxiety Inventory (BAI; Beck et al., 1988); (6) Beck Depression Inventory (BDI; Beck et al., 1979); (7) Substance Abuse Subtle Screening Inventory (SASSI–2), an 88-item questionnaire that sampled a broad range of alcohol and drug abuse patterns, including subtle attributes of abuse and denial (Miller, 1988); (8) Symptom CheckList (SCL–90–R), a 90-item questionnaire that provided a direct assessment of psychological symptoms and a limited range of health disorders (Derogatis, 1994; Derogatis et al., 1974, 1976); (9) Minnesota Multiphasic Personality Inventory–2 (MMPI–2), the first 370 items (basic form; Graham, 1993; Hathaway et al., 1989); (10) The Positive Affect/Negative Affect Schedule (PANAS), a 20-item questionnaire that assessed the emotional style a person uses to cope with life–world events (Watson et al., 1988); (11) Life Experience Scale (LES), a 57-item questionnaire that provided a current measure of present-life stressors that may color the presentation of symptoms (Sarason et al., 1978); and (12) Combat Exposure Scale (CES), a 50-item questionnaire that assessed war-zone exposure to violence, wounding, wounding–death of others, threat of severe injury–death, leadership failures, abusive violence, and POW captivity (Keane et al., 1989; Wolfe et al., 1993). The LES and CES are not included in the results but were identified here to fully describe the methods used.

The published scales were evaluated for each of the 10 psychological tests in the Results. However, for reporting purpose, scales were selected for presentation from the tests with multiple scales in order to avoid correcting statistically for comparisons of scales collateral to our main interest. The following scale selections were based on our expectations of important factors: (1) the Physical Functioning and Health Perception scales of the SF36 (as measures of self-perception of function and health); (2) the SASSI–2 Obvious Attributes (OAT) scale (because this scale samples direct questions about substance use rather than hypothetical correlates and because OAT test–retest reliability data were available in the SASSI manual); (3) the SCL–90–R global severity index or GSI scale (because it consists of all test items); (4) the HS, D, and HY scales from the MMPI–2 (because elevations of HS and HY are associated with psychogenic origin of somatic symptoms; Hathaway et al., 1989); (5) the Negative Influencing Factors (NF) scale of the PANAS (because it is more sensitive to psychopathology and has been recommended by the test's authors as a standalone screening instrument; Watson et al., 1988).

Neurobehavioral Performance Tests

The six BARS neurobehavioral tests, were as follows: (1) Simple Reaction Time, a test of response speed (e.g., Posner, 1978); the mean reaction time was calculated from all trials; (2) Selective Attention Test (SAT), which assesses attention by presenting dots (approximately 1 mm diameter) on a variable interstimulus interval schedule inside or outside two squares and requiring a differential response (*3* or *7* button) for dots inside the left or right square, respectively, and withholding of a response for dots outside the squares (Anger et al., 1996); the mean percent correct was measured; (3) Digit Span (Wechsler, 1955); the maximum span forward and reverse (the test terminated on two errors at the same span length) was the measure from this test; (4) Symbol Digit, in which symbols were paired in a matrix with numbers as in a code and the respondent typed into a matrix of symbols but no numbers, the number associated with the symbol (Smith, 1968); time to complete five matrices was analyzed; (5) Serial Digit Learning, employing a nine-digit series and a maximum of 12 trials (Benton et al., 1994); score ranging from zero to 24 calculated with 2 points for each correct trial and 1 point for partial correct trials; and, (6) Oregon Dual Task Procedure (ODTP), a new im-

https://doi.org/10.1017/S1355617799533031 Published online by Cambridge University Press

plementation of a test of motivation, attention and memory. The ODTP consisted of the presentation of a five-digit number; followed by a vigilance (distractor) task for 5-, 15-, or 25-s intervals; followed by a forced choice between the original five-digit number and a different (incorrect) alternative. The number correct and mean latency on the digit memory trials and the mean latency for correct trials on the distractor task were measured; the median latency is presented in the results. This test was designed to be parallel to the Portland Digit Recognition Test (PDRT) which has been used to measure motivation to perform poorly on neurobehavioral tests (Binder, 1993; Binder & Kelly, 1996; Binder & Willis, 1991). The individually administered PDRT chiefly differs from the ODTP by employing backward counting as the distractor task, and it does not include a measure of the latency between the appearance of the forced choice option and the participant response.

The test–retest reliability (1 week between tests) of all tests in this study given in the same order was examined in 45 volunteer *test–retest referents* recruited from the general Portland, OR, population by a newspaper advertisement. The reliabilities were within the range reported for the tests in their original formats (Campbell et al., 1999; Rohlman et al., 1997).

RESULTS

Recruitment and case disposition were ongoing activities. Following the exclusion of questionnaire respondents who (1) failed to meet the case definition during the phone followup (e.g., recanted the time of symptom development as during or after the PGW); (2) were ineligible (e.g., Viet Nam war veteran); (3) reported exclusionary medical diagnoses (e.g., had diabetes); or (4) declined to participate in further testing, 152 respondents were successfully recruited for testing. Of these 152, case disposition was decided for 101 as of December, 1996 (the approximate midpoint of our study); they are the subject of this report. Of the 101, 65% met study criteria for *cases* and 35% met the criteria for *controls*. This achieved the goal of the PEHRC case definition to encompass a very broad diversity in symptoms and allow analytic techniques to reveal unique constellations of symptoms or health outcomes against a control background in which those symptoms were completely absent.

As seen in Table 1, cases and controls did not differ with respect to age; years of education; or, when obtainable, mental aptitude scores on the Armed Forces Qualifying Test (AFQT), a self-administered test taken by all military personnel at induction (Welsh et al., 1990). Study participants were both male and female, had served in all military branches, and included veterans who were on active or activated reserve status at the time of deployment. Approximately 15% of our study participants reported taking examinations offered by the DVA or DOD as part of a PG Registry (CRHCSPGW, 1996; Newmark and Clayton, 1995), which approximates the proportion of Oregon veterans in the Portland PG Registry (for the series reported here, mailings were primarily sent to Portland-area veterans).

Psychological and Neurobehavioral Tests

Cases differed from controls on virtually all psychological (HSS) test scales in the direction of increased distress. These differences remained significant in the 13 scales selected for analysis [$t(98–101) = 16.61–60.53$, $ps \leq .0001$], even after a stringent Bonferroni correction (for 13 comparisons, $p \leq .0038$) for multiple comparisons. Virtually every scale not selected for presentation reveals the same substantial differences between cases and controls as the selected scales. Case performance was inferior to control performance on all neurobehavioral tests, and measures on three of the six tests were statistically significant ($p \leq .05$). However, the only difference that remained significant after a Bonferroni correction (for six comparisons, $p \leq .0056$) was the forced-choice latency of the Oregon Dual Task procedure (ODTP), which measures memory for digits after a delay interval [$t(99) = 11.0199$, $p < .0013$].

Inspection of the individual ODTP forced-choice latencies does not reveal a distribution-wide shift of the case group. Rather, the majority of cases have scores that lie centrally within the range of control scores, but a smaller group of cases lie outside the range of virtually all control scores (Figure 1). It is the latter subgroup that is responsible for the statistically significant differences seen between cases and controls on the ODTP. The bottom panel of Figure 1 depicts the ODTP forced-choice latencies from the 45 test–retest reliability referents (Campbell et al., 1999; Rohlman et al., 1997). They suggest that the control group distribution is normal and that the control distribution may also be representative of the U.S. population that did not serve in the Gulf. Parenthetically, the test–retest referents from the general population were not given medical screen-

Table 1. Demographic variables for cases and controls

		Age (years)		Education (years)		AFQT[a]		Percent	Percent
Group	*N*	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	male	employed[b]
Cases	66	32.6	(7.8)	13.5	(1.6)	57.7	(20.0)	80	94
Controls	35	30.6	(7.3)	13.8	(2.2)	64.3	(20.1)	83	94

[a]AFQT $N = 35$ cases and 20 controls. [b]Percent employed full or part time.

https://doi.org/10.1017/S1355617799533031 Published online by Cambridge University Press

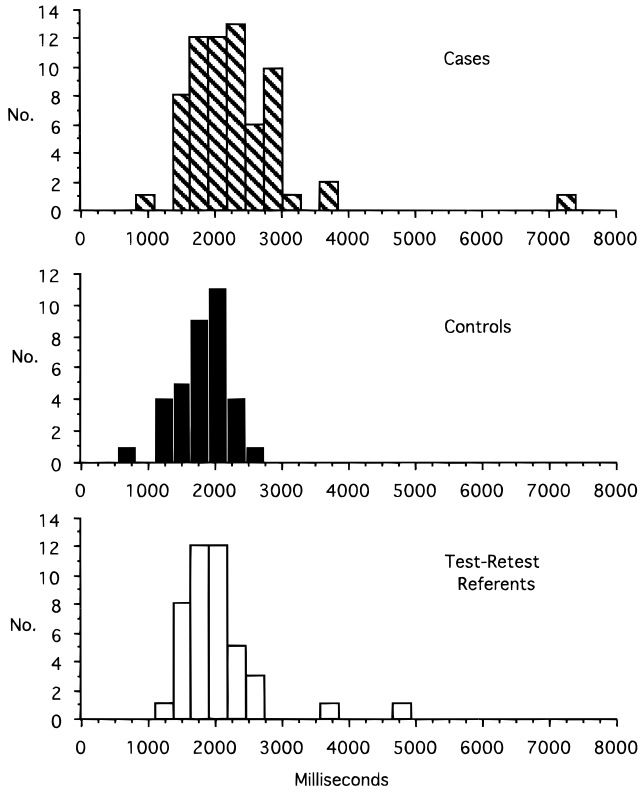

Fig. 1. Frequency histograms of mean ODTP forced-choice latencies in cases ($N = 35$), controls ($N = 66$) and test–retest referents ($N = 43$; see Methods).

ing examinations to exclude those with medical conditions as in the study described here.

Identification of a "Slow Case" Subgroup

The bimodal appearance of the distribution of ODTP forced-choice latency scores (Figure 1) suggested that cases might consist of two distinct distributions. To test this hypothesis, cases were divided into *slow case* and *other case* subgroups at an arbitrary cut-off score 2 standard deviations above (slower than) the mean ODTP forced-choice latency of the control group. The resulting slow case subgroup (of 13 participants) has a mean ODTP latency (2.7 s; $SD = 0.19$) that is approximately 3 standard deviations slower than the control mean (1.7 s; $SD = 0.37$). The mean was calculated from the median of each participant's ODTP forced-choice latency distribution to eliminate any impact of the extreme case outlier seen in Figure 1. Four-group ANOVA [$F(3,140) = 23.99$, $p \leq .0001$] followed by Bonferroni-adjusted *post-hoc* comparisons supports the statistical significance ($p \leq .05$) of the ODTP latency differences between the slow cases and the other cases, controls, and test–retest referents.

Slow Cases: Psychological *Versus* Neurobehavioral Test Results

The hypothesis that the slow cases constituted a unique subgroup was further tested by a comparison of their test results with the remaining other cases and controls. On the 13 selected psychological test measures, both the slow case and other case subgroups had significantly higher levels of abnormal responses when compared to controls ($p \leq .003$; all probabilities are below a Bonferroni-adjusted .05 cut-off of $p = .0036$). This suggests that slow cases and other cases share the same psychological distress. In contrast, other case performance was similar to control performance on the neurobehavioral tests, while slow cases were significantly slower (below a Bonferroni-adjusted cut-off $p = .0056$) than the controls on Symbol Digit [$F(2,98) = 7.26$, $p = .0012$], Simple Reaction Time [$F(2,97) = 10.08$, $p \leq .0001$], Digit Span Forward [$F(2,91) = 6.42$, $p = .0026$], and Digit Span Backward [$F(2,88) = 6.78$, $p = .0019$], although not on the Selective Attention [$F(2,96) = 3.02$, $p = .054$] or Serial Digit Learning tests [$F(2,78) = 1.76$, $p = .179$; Table 2].

Parenthetically, the individual in the slow case subgroup who provided the extreme outlier data point in the ODTP forced-choice latency distribution (Figure 1, top panel) was not at the extremes of the slow case distributions on the remaining neurobehavioral or psychological tests and was therefore included in all inferential analyses in this article except those involving the ODTP. Some of the analyses were repeated with that individual excluded to assess the impact empirically. The inferential results reported above did not differ substantially with or without that individual.

Slow Cases: Premorbid Ability

The possibility that the slow cases had lower preservice functional capacity than other cases or controls was suggested by the group differences on the AFQT scores (Table 2), although the differences were not significant. Since AFQT data were available for only 55% of our participant sample, their value is minimized. To assess the possible impact of premorbid ability reflected by AFQT results, AFQT and Simple Reaction Time scores from cases and controls were entered into a multiple regression equation predicting ODTP latency. Forcing AFQT scores into the regression predicted only 11% of the variance [$F(1,68) = 8.77$, $p \leq .004$]. Adding Simple Reaction Time as a variable to AFQT predicted 31% of the variance in AFQT scores [$F(2,67) = 15.16$, $p < .0001$], suggesting that premorbid ability was not the primary explanation for the ODTP results in slow cases, although it has some influence on those results.

Slow Cases: Motivation

ODTP performance provided suggestive information on motivation, a critical variable on neurobehavioral tests. The ODTP was developed as a computer-implemented analog of the PDRT. Poor motivation is inferred by performance below 54% correct on the forced-choice portion of the PDRT, based on the probability of guessing correctly on each item (.5) and empirical data obtained from patients with well-documented dysfunction and those seeking financial compensation for mild head trauma (but with no neurologic

https://doi.org/10.1017/S1355617799533031 Published online by Cambridge University Press

Table 2. Means (*SD*s) and ANOVA comparison *p*s in controls, other cases, and slow cases for psychological and neurobehavioral tests

	C		OC		Slow		C *vs.* OC	C *vs.* Slow	OC *vs.* Slow
Test/Measure	*M*	(*SD*)	*M*	(*SD*)	*M*	(*SD*)	*p*	*p*	*p*
Demographics									
Age	30.6	(7.3)	32.3	(7.5)	34.2	(8.9)	ns	ns	ns
Education	13.8	(2.2)	13.5	(1.5)	13.8	(1.7)	ns	ns	ns
Armed Forces Qualifying Test (AFQT)	64.3	(20.1)	60.4	(19.0)	50.1	(22.0)	ns	ns	ns
Days in Persian Gulf (per DOD records)	120.7	(45.9)	117.8	(60.9)	129.6	(77.0)	ns	ns	ns
Posttraumatic stress disorder									
Mississippi	51.5	(10.7)	74.0	(16.0)	74.9	(19.6)	*	*	ns
Penn	13.2	(5.9)	27.3	(10.4)	27.5	(9.9)	*	*	ns
PCL–C	20.3	(5.3)	32.5	(9.9)	32.8	(10.7)	*	*	ns
Self-reported health									
SF-36–health perception	86.9	(12.1)	55.3	(20.8)	65.2	(19.6)	*	*	ns
SF-36–physical functioning	98.3	(3.8)	81.2	(16.5)	85.8	(17.9)	*	*	ns
Symptoms of psychopathology									
Beck Anxiety Inventory	2.3	(3.2)	10.9	(7.1)	11.5	(7.0)	*	*	ns
Beck Depression Inventory	4.8	(3.8)	13.6	(6.7)	12.9	(5.9)	*	*	ns
SASSI–2–obvious attributes	4.7	(2.6)	7.1	(2.6)	6.2	(2.7)	*	ns	ns
SCL–90–R–global severity	0.24	(0.25)	0.88	(0.52)	0.86	(0.44)	*	*	ns
Personality measures									
MMPI–2–HS	48.7	(8.4)	66.8	(13.3)	60.7	(15.6)	*	*	ns
MMPI–2–D	47.3	(6.9)	64.5	(13.4)	64.6	(14.0)	*	*	ns
MMPI–2–HY	48.6	(7.3)	61.2	(12.8)	57.7	(13.0)	*	*	ns
PANAS–negative factors	14.3	(4.1)	20.8	(6.4)	20.2	(5.0)	*	*	ns
Neurobehavioral tests									
Simple Reaction Time	314	(56)	314	(48)	406	(139)	ns	**	**
Selective Attention Test–percent correct	94	(6)	92	(9)	85	(24)	ns	**	ns
Digit Span Forward	6.7	(1.7)	6.2	(1.3)	4.9	(1.1)	ns	**	**
Digit Span Backward	6.4	(2.1)	6.1	(1.7)	4.2	(1.0)	ns	**	**
Symbol Digit–latency	1643	(364)	1741	(402)	2106	(299)	ns	**	**
Serial Digit Learning	18.6	(4.7)	16.1	(6.4)	15.1	(4.1)	ns	ns	ns
ODTP–correct (out of 48 possible)	47.8	(0.5)	47.2	(1.6)	45.6	(3.2)	ns	**	**
ODTP–forced choice latency (slow case outlier removed)	1708	(375)	1846	(320)	2688	(186)	ns	**	**
ODTP–distractor task latency	543	(72)	574	(107)	631	(108)	ns	ns	ns

Note. C = controls; OC = other cases; Slow = slow cases. The *N* for most comparisons was 35 controls, 53 other cases, and 13 slow cases (less 1 in some cells due to data loss); the *N* was 20, 26, and 9 for the AFQT data; 20, 50, and 11 for Serial Digit Learning; 29, 53, and 12 for the Digit Span Forward Test; and 29, 50, and 11 for the Digit Span Backward test, respectively for controls, other cases and slow cases.
*$p \leq .0033$; **$p \leq .0056$.

evidence of brain dysfunction; Binder, 1993; Binder & Kelly, 1996). However, in the present study, ODTP forced-choice performance by all participants was better than 80%. Thus, there was no indication of poor motivation on this task, supporting by inference that participants were motivated to perform well on the neurobehavioral tests.

Slow Cases: ODTP Correlations With Other Tests

The ODTP forced-choice latency score was not strongly related to the psychological measures ($r < .20$ on all psychological test measures in Table 1 except the MMPI–2 D scale which had a correlation of .24, $p = .02$). This contrasts with moderate correlations between ODTP latency and the neurobehavioral tests (rs = .24–.28 with Serial Digit Learning and SAT; rs = .35–.43 with Digit Span Forward and Backward and Symbol Digit; and $r = .47$ with Simple Reaction Time; ps $< .03$–.001). Thus, ODTP forced-choice latency appears to be related to the other neurobehavioral performance scores and not to abnormal psychological function. ODTP forced choice latency correlated minimally with age ($r = .20$; $p = .04$) and was not significantly related to education ($r = .05$; $p = .62$).

Slow Cases: Demographic Factors

The demographic characteristics of the 13 slow cases were not remarkable when compared with controls and other cases (Table 2), including the mean age (slow cases were 3 years older than the controls) and education. Additional demographic data (Table 3) further support the comparability with the controls, although some differences emerge (chi-square tests). Most slow cases (8 of 13) reported being married or

https://doi.org/10.1017/S1355617799533031 Published online by Cambridge University Press

Table 3. Percent controls, other cases and slow cases making positive questionnaire responses, and significance of chi-square test: Supplementary demographics

Question	Control	Other case	Slow case	*p*
Number	35	53	13	–
Percent male	80	87	54	.03
White	94	96	92	ns
Marital status				ns
Married–Significant other	49	64	62	
Divorced–separated	20	15	15	
Never married	29	21	23	
Years of education completed				ns
<12	3	8	0	
12	29	25	25	
13–16	57	61	62	
>16	12	8	16	
Service branch				ns
Army	20	34	46	
Air Force	11	8	0	
Navy	34	23	15	
Marines	23	21	23	
National Guard	11	23	15	
Coast Guard	0	2	0	
Promoted during PG service	11	13	23	ns
Employed full or part time	89	92	100	ns
Income status				ns
Can't ever make ends meet	9	9	0	
Sometimes can't make ends meet	9	23	23	
Just enough, no more	23	26	15	
Some months have money left over	45	32	54	
Always have money left over	11	9	8	
Reported cognitive–psychological symptoms	0	76	77	<.0001
Sought medical attention in PG	29	57	77	.01
Took prescription medications in PG	11	30	38	.03

Note. The number of years of education completed was filled in by each respondent, and the chi-square test was run on the uncombined data; the data are combined in the table to provide a more digestible presentation. Answers that do not add to 100% lack responses that were not recorded; rounding caused some answers to total more than 100%. No statistical correction for multiple comparisons has been applied to the chi-square analyses; over 190 discrete items were on the questionnaire.

living with a significant other, and all but one slow case was White, mirroring Oregon's low percentage of minorities. At the time of the PG War, slow cases were members of four branches of the military. Eleven of the 13 slow cases reported being employed full time and the other 2 part time when they were tested, suggesting that they were productive members of the society to which they returned after the war, and their income status further supports this contention (Table 3).

The most noteworthy demographic factor (Table 3) is that a higher percentage of slow cases were female (46%) than in the other case (13%) or control (20%) groups [$\chi^2(4, N = 101) = 10.5$, $p = .03$]. The mean slow case ODTP forced-choice latency of females was slower than that of slow case males, but not significantly so with the outlier in Figure 1 [$t(11) = 137$, $p = .197$, 2-tailed] nor without the outlier [$t(10) = 1.71$, $p = .117$, 2-tailed]. Further, the ODTP latency distribution of test–retest referents (see Methods), a group in which women also comprise half of the sample, approximates that of the controls, not the slow cases (Figure 1). Finally, our study intentionally oversampled women to the increase sample size of the study population, and sampled them early in the study.

The questionnaire filled out by all PEHRC study participants provides information about factors that might explain the neurobehavioral performance. The approximately 200 questions related to exposure variables did not reveal any single obvious factor distinguishing slow cases from the other participants, although analysis of the complete sample may reveal important factors only suggested at this point. A higher percentage of slow cases reported seeking medical treatment (77%) and taking prescription medicine (38%) during their service in the Gulf when compared to the other cases (57% and 30%, respectively) or controls (29% and 11%); the group differences were significant [$\chi^2(4, N = 101) = 13.2$, $p = .01$; and $\chi^2(4, N = 101) = 10.5$, $p = .03$, respectively] for seeking medical treatment and taking prescription medicine. This provides support for the hypothesis that the slow cases could have fallen in the unhealthy end of the PG population. On the other hand, 3 of the slow cases did

https://doi.org/10.1017/S1355617799533031 Published online by Cambridge University Press

not report cognitive or affective symptoms (but were classified as cases by endorsing other symptoms), and only 1 slow case (8%) reported that they had enrolled in the PG Registry (compared with 15% in our overall study sample). No statistical correction has been applied to the chi-square analyses reported here; over 200 discrete items were on the questionnaire.

DISCUSSION

Our results reveal that Persian Gulf veterans reporting PG symptoms have slower neurobehavioral performance on a test of digit recall and that they reveal increased distress on a broad range of psychological measures, compared to Persian Gulf veterans who do not report such symptoms. These results from a sample of the general PG War veteran population are more consistent with Axelrod and Milner (1997) who reported small Stroop and Grooved Pegboard performance deficits and Goldstein et al. (1996), who reported a significant impairment in a summary index in their PG group, than they are with the conclusions of Haley et al. (1997), Haley and Kurt (1997), and Hom et al. (1997) that there are extensive neuropsychological differences indicative of toxic encephalopathy in symptomatic PG veterans. It may be that the Hom and Haley findings are representative of older (*M* age = 47.9 years; *M* education = 11.9 years) veterans in the tail of the distribution of illness in their battalion. Further, statistical corrections for multiple comparisons were not applied to the 70+ dependent variables identified by Hom and Haley, who asserted that the symptomatic veteran group had lower scores on 59 of the measures.

Our data, however, do not support a distribution-wide neurobehavioral deficit in veterans reporting PG symptoms. Rather it appears that a small number of symptomatic veterans respond very slowly on a recall test, and that other tests measuring response speed reveal comparable differences in this group we have designated as slow cases. The marked differences between slow cases and other cases on the neurobehavioral tests (but not on the psychological tests) supports the hypothesis that the slow cases constitute a unique subgroup that is distinct from the other cases as well as the controls. Slow cases had objective deficits, compared to control performance, in measures of memory (likely including a substantial working memory component), attention, and response speed revealed by neurobehavioral tests. The similarity between slow cases and other cases on the psychological tests suggests that the slow cases are not the tail of the distribution of psychological dysfunction. Emotional functioning endorsed by Sillanpaa et al. (1997) as the most influential factor in symptom development does not appear to be the primary factor in explaining the results of the slow cases on the neurobehavioral performance tests in our sample. In fact, compared to referents, the significant performance deficits seen in the slow cases are consistent with the deficits reported in organophosphate-poisoned workers (viz., Symbol Digit, Digit Span, vigilance tests; Rosenstock et al., 1991; Savage et al., 1988; Steenland et al., 1994). While this is also consistent with the Haley et al. (1997), Haley and Kurt (1997), and Hom et al. (1997) conclusions, we have no evidence to support their proposal that toxic substances encountered in the PG are associated with the deficits in our more diverse sample.

Sample Characteristics

The results described here are based on a self-selected sample of veterans who responded to a population-based mail questionnaire. The sample who received the mail questionnaire was randomly selected from all veterans from Oregon and Southwest Washington who were in the DOD database of Desert Storm–Desert Shield veterans. The questionnaire return rate over 60% builds confidence in the generalizibility of the survey data and exceeds the response rates frequently obtained in mail surveys (Eaker et al., 1998). Our study is not without bias, however. We do not have any health information on the 38% of veterans who did not return the questionnaires. Furthermore, our case–control study was based on veterans who volunteered (for $50) to travel to the Portland VA Medical Center and participate in 8 hr of testing and therefore may not be representative of all veterans returning the surveys. However, our selection procedures produced a diverse sample with representation from all branches of U.S. military service and many major work–duty classifications (not described here). Further, only 15% of our sample had sought and received PG Registry examinations. These factors are strongly suggestive that the study reflects a population-based sample. In contrast, other published reports have focused on samples drawn from either a PG Registry or military units with a high percentage of complaints.

Of those questionnaire respondents who declined testing, the most frequent reason given was work responsibilities. No veterans refused to participate due to the severity of their symptoms (McCauley et al., 1997). Nevertheless, the voluntary nature of the case–control study necessarily affects the generalizibility of the findings to an unknown degree (Ellenberg, 1994). It is quite possible that veterans worried about their health are more heavily represented in our sample than those who were not worried about their health. While such bias would affect prevalence issues (not addressed for that reason), the nature of dysfunction or deficit in cases can be identified as long as a sample of asymptomatic veterans serves as controls. This was the case here.

Implications of the Slow Case Subgroup

The slow cases described here provide objective evidence of the existence of one constellation of PG symptoms reported by a substantial percentage of returning veterans. Potentially, the size of this subgroup could approach 10% of Gulf War veterans, based on the percentage with this deficit in our population-based study (this further assumes conservatively that ODTP performance by the 38% of veterans who did not return questionnaires would be comparable to controls). Furthermore, these slow cases offer the potential

https://doi.org/10.1017/S1355617799533031 Published online by Cambridge University Press

for discerning the etiology of cognitive complaints for which this study provides objective evidence in a subgroup of cases. The etiology could involve exposure to neurotoxic substances such as organophosphates in the Gulf, but there are other equally plausible explanations that have no relationship to service in the Gulf (e.g., premilitary differences in health, lifestyle, hobbies, former or current occupation). Focused examination and in-depth analysis of this subgroup could reveal an important, currently unrecognized factor that is associated with service in the Gulf, although that service may be only coincidentally related to those deficits. Finally, while psychological distress may be highly associated with symptoms in some veterans, concurrent neurobehavioral effects may also be detectable in some veterans. Thus, discovery of psychological distress should not distract the clinician from pursuit of neurobehavioral deficits.

There is no evidence that the PG symptoms represent a single unitary disorder, and indeed it seems unlikely. While our small sample of slow cases appears to have a unique constellation of neurobehavioral deficits, compared to controls, it may be that the difference in ODTP performance eventually will be found to be spurious. Perhaps the important point to emerge from this result is that the broad range of reported PG symptoms may obscure individual problems in small groups of veterans and that only by extracting those small groups from the larger background symptom noise will important factors emerge. More succinctly, large group studies may reveal critical subgroups on close inspection of their results.

ACKNOWLEDGMENTS

This research was supported by the Department of Veterans Affairs and the Portland Environmental Hazards Research Center, a joint project of the Portland Veterans Affairs Medical Center and the Center for Research on Occupational and Environmental Toxicology of the Oregon Health Sciences University. Appreciation is extended to persons contributing to critical aspects of this project, including Nadia Redmond; Michael Lasarev; Sandy Joos, Ph.D.; Wendy Johnston, M.D.; Michael Wynn, M.D.; Andre Barkhuizen, M.D. Center oversight and more direct project contributions (e.g., to study design, case definitions) were provided by Peter S. Spencer, Ph.D., and Dennis Bourdette, M.D. Dr. Daniel Storzbach made the critical contribution of recognizing the bimodality of the *slow case* distribution and characterizing this subgroup.

REFERENCES

Anger, W.K. (1990). Worksite behavioral research: Results, sensitive methods, test batteries, and the transition from laboratory data to human health. *Neurotoxicology*, *11*, 629–720.

Anger, W.K., Rohlman, D.S., & Sizemore, O.J. (1994). A comparison of instruction formats for administering a computerized behavioral test. *Behavior Research Methods, Instruments, & Computers*, *26*, 209–212.

Anger, W.K., Rohlman, D.S., Sizemore, O.J., Kovera, C.A., Gibertini, M., & Ger, J. (1996). Human behavioral assessment in neurotoxicology: Producing appropriate test performance with written and shaping instructions. *Neurotoxicology and Teratology*, *18*, 371–379.

Axelrod, B.N. & Milner, I.B. (1997). Neuropsychological findings in a sample of Operation Desert Storm veterans. *Journal of Neuropsychiatry*, *9*, 23–28.

Beck, A.T., Epstein, N., Brown, G., & Steer, R.A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology*, *56*, 893–897.

Beck, A.T., Rush, A.J., Shaw, B.F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford.

Benton, A.L., Sivan, A.B., Hamsher, K. D. S., Varney, N.R., & Spreen, O. (1994). *Contributions to neuropsychological assessment*. New York: Oxford University Press.

Binder, L.M. (1993). Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. *Journal of Clinical and Experimental Neuropsychology*, *15*, 170–182.

Binder L.M. & Kelly, M.P. (1996). Portland digit recognition test performance by brain dysfunction patients without financial incentives. *Assessment*, *3*, 403–409.

Binder, L.M. & Willis, S.C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment*, *3*, 175–181.

Butcher, J.N., Dahlstrom, W.G., Graham, J.R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2: Manual for administration and scoring*. Minneapolis: University of Minnesota Press.

Campbell, K.A., Rohlman, D.S., Storzbach, D., Binder, L.M., Anger, W.K., Kovera C.A., Davis, K.L., & Grossmann, S. (1999). Test-retest reliability of psychological and neurobehavioral tests self-administered by computer. *Assessment*, *6*, 21–32.

Centers for Disease Control. (1995). Unexplained illness among Persian Gulf War veterans in an Air National Guard unit: Preliminary report—August 1990–March 1995. *Morbidity and Mortality Weekly Report*, *44*, 443–447.

Committee to Review the Health Consequences of Service During the Persian Gulf War. (1996). In *Health consequences of service during the Persian Gulf War: Recommendations for research and information systems*. Washington, DC: U.S. Institute of Medicine, National Academy Press.

Derogatis, L.R. (1994). *Symptom Checklist–90–R (SCL–90–R). Administration, scoring, and procedures manual* (3rd ed.). Minneapolis, MN: National Computer Systems, Inc.

Derogatis, L.R., Lipman, R.S., Rickels, K., Uhlenhuth, E.H., & Covi, L. (1974). The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Behavioral Science*, *19*, 1–15.

Derogatis, L.R., Rickels, K., & Rock, A. (1976). The SCL–90 and the MMPI: A step in the validation of a new self-report scale. *British Journal of Psychiatry*, *128*, 280–289.

Dikmen, S.S., Reitan, R.M., Temkin, N.R., & Machamer, J.E. (1992). Minor and severe head injury: Emotional sequelae. *Brain Injury*, *6*, 477–478.

Eaker, S., Bergstrom, R., Bergstrom, A., Adami, H-O., & Nyren, O. (1998). Response rate to mailed epidemiologic questionnaires: A population-based randomized trial of variations in design and mailing routines. *American Journal of Epidemiology*, *147*, 74–82.

Ellenberg, J.H. (1994). Cohort studies: Selection bias in observational and experimental studies. *Statistics in Medicine*, *13*, 557–567.

Goldstein, G., Beers, S.R., Morrow, L.A., Shemansky, W.J., & Stein-

hauer, S.R. (1996). A preliminary neuropsychological study of Persian Gulf Veterans. *Journal of the International Neuropsychological Society*, *2*, 368–371.

Graham, J.R. (1993). *MMPI–2: Assessing personality and psychopathology* (2nd ed.). New York: Oxford University Press.

Haley, R.W., Hom, J., Roland, P.S., Bryan, W.W., Van Ness, P.C., Bonte, F.J., Devous, M.D., Mathews, D., Fleckenstein, J.L., Wians, F.H., Wolfe, G.I., & Kurt, T.L. (1997). Evaluation of neurologic function in Gulf War veterans. A blinded case–control study. *Journal of the American Medical Association*, *277*, 223–230.

Haley, R.W. & Kurt, T.L. (1997). Self-reported exposure to neurotoxic chemical combinations in the Gulf War: A cross-sectional epidemiologic study. *Journal of the American Medical Association*, *277*, 231–237.

Hammarberg, M. (1992). Penn Inventory for Posttraumatic Stress Disorder: Psychometric properties. *Psychological Assessment*, *4*, 67–76.

Hathaway, S.R., McKinley, J.C., & the MMPI Restandardization Committee of the University of Minnesota Press: Butcher, J.N., Dahlstrom, W.G., Graham, J.R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory–2: Manual for administration and scoring*. Minneapolis, MN: University of Minnesota Press.

Hom, J., Haley, R.W., & Kurt, T.L. (1997). Neuropsychological correlates of Gulf War Syndrome. *Archives of Clinical Neuropsychology*, *12*, 531–544.

Hyams, K.C., Wignall, F.S., & Roswell, R. (1996). War syndromes and their evaluation: From the U.S. Civil War to the Persian Gulf War. *Annals of Internal Medicine*, *125*, 398–405.

Iowa Persian Gulf Study Group. (1997). Self-reported illness and health status among Gulf War veterans. A population-based study. *Journal of the American Medical Association*, *277*, 238–245.

Keane, T., Caddell, J., & Taylor, K.L. (1988). Mississippi scale for combat-related posttraumatic stress disorder: Three studies in reliability and validity. *Journal of Consulting and Clinical Psychology*, *56*, 85–90.

Keane, T.M., Fairbank, J.A., Caddell, J.M., Zimering, R.T., Taylor, K.L., & Mora, C.M. (1989). Clinical evaluation of a measure to assess combat exposure. *Psychological Assessment*, *1*, 53–55.

Kovera, C.A., Anger, W.K., Campbell, K.A., Binder, L.M., Storzbach, D., Davis, K.L., & Rohlman, D.S. (1996). Computer-administration of questionnaires: A health screening system (HSS) developed for veterans. *Neurotoxicology and Teratology*, *18*, 511–518.

McCauley, L. Joos, S., Bourdette, D., Johnston, W., Barkhuizen, A., Wynn, M., & Redmond, N. (1997, November). *Persian Gulf War illnesses: Validation of self-reported symptoms by clinical examination*. Paper presented at the meeting of the American Public Health Association, Indianapolis, IN.

Miller G.A. (1988). *Substance Abuse Subtle Screening Inventory (SASSI) manual*, Spencer, IN: Spencer Evening World.

Newmark, J. & Clayton III, W.L. (1995). Persian Gulf illnesses: Preliminary neurological impressions. *Military Medicine*, *160*, 505–507.

Persian Gulf Veterans Coordinating Board. (1995). Unexplained illnesses among Desert Storm veterans. A search for causes, treatment, and cooperation. *Archives of Internal Medicine*, *155*, 262–268.

Posner, M.I. (1978). *Chronometric explorations of mind*. Hillside, NJ: Lawrence Erlbaum Associates.

Rohlman, D.S., Anger, W.K., Binder, L.M., Campbell, K.A., Kovera, C.A., Storzbach, D, Davis, K.L., & Faring-Grund, L. (1997, August). Linked behavioral and psychological assessments of Persian Gulf veterans. In R.F. White (Chair), *Persian Gulf War Veterans' Health Complaints—Psychological Studies*. Symposium conducted at the meeting of the American Psychological Association, Chicago, IL.

Rohlman, D.S., Sizemore, O.J., Anger, W.K., & Kovera, C.A. (1996). Computerized neurobehavioral testing: Techniques for improving test instructions. *Neurotoxicology and Teratology*, *18*, 407–412.

Rosenstock, L., Keifer, M., Daniell, W.E., McConnell, R., & Claypoole, K. (1991). Chronic central nervous system effects of acute organophosphate pesticide intoxication. *Lancet*, *338*, 223–227.

Ryan, C.M., Morrow, L.A., Bromet, E.F., & Parkinson, D.K. (1987). Assessment of neurological dysfunction in the workplace: Normative data from the Pittsburgh Occupational Exposures test battery. *Journal of Clinical and Experimental Neuropsychology*, *9*, 665–679.

Sarason, I.G., Johnson, J.H., & Siegel, J.M. (1978). Assessing the impact of life changes: Development of the Life Experiences Survey. *Journal of Consulting and Clinical Psychology*, *46*, 853–863.

Savage, E.P., Keefe, T.J., Mounce, L.M., Heaton, R.K., Lewis, J.A., & Burcar, P.J. (1988). Chronic neurological sequelae of acute organophosphate pesticide poisoning. *Archives of Environmental Health*, *43*, 38–45.

Sillanpaa, M.C., Agar, L.M., Milner, I.B., Podany, E.C., Axelrod, B.N., & Brown, G.G. (1997). Gulf War veterans: A neuropsychological examination. *Journal of Clinical and Experimental Neuropsychology*, *19*, 211–219.

Smith, A. (1968). SDMT: A neuropsychological test for economic screening. *Hearing Disorders*, *3*, 83–91.

Steenland, K., Jenkins, B., Ames, R.G., O'Malley, M., Chrislip, D., & Russo, J. (1994). Chronic neurological sequelae to organophosphate pesticide poisoning. *American Journal of Public Health*, *84*, 731–736.

Ware, J.E. & Sherbourne, C.D. (1992). The MOS 36-item Short-Form Health Survey (SF-36): 1. Conceptual framework and item selection. *Medical Care*, *30*, 473–483.

Ware, J.E., Sherbourne, C.A., & Davies, A.R. (1988). *A short-form general health survey*. Santa Monica, CA: RAND Corporation, Publication No. P-7444.

Watson, D., Clark, L.A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070.

Weathers, F.W., Litz, B.T., Herman, D.S., Huska, J.A., & Keane, T.M. (1993, October). *The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility*. Paper presented at the meeting of the International Society for Traumatic Stress Studies, San Antonio, TX.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale*. New York: The Psychological Corporation.

Welsh, J.R., Kucinkas, S.K., & Curran, L.T. (1990). *Armed Services Vocational Battery (ASVAB): Integrative review of validity studies*. Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Wolfe, J., Brown, P.J., Furey, J., & Levin, K.B. (1993). Development of a wartime stressor scale for women. *Psychological Assessment*, *5*, 330–335.