# Evaluation of Neuromuscular Symptoms in UK Gulf War Veterans. A Controlled Study

[1,2] M.R. Rose, MD; [1,2] M.K. Sharief, MD PhD; [1] J. Priddin, BSc; [1]V. Nikolaou, MSc; [1]L. Hull, BSc; [1]C. Unwin, MSc; [3]R. Ajmal-Ali, MSc; [3]R.A. Sherwood, M.Sc, D.Phil [1]A.Spellman, BSc; [1]A. David, MD; [1]S. Wessely, MD

From

1) The Gulf War Illness Research Unit; Guy's, King's & St Thomas' School of Medicine, London, England

2) The Department of Neurology; Guy's, King's & St Thomas' School of Medicine, and King's College Hospital, London, England

3) The Department of Clinical Biochemistry, King's College Hospital, London England

**Corresponding Author:**

M.R. Rose, Department of Neurology, King's College Hospital, London SE5 9RS, England

Tel +44 20 7346 5355

Fax +44 20 7346 5329

E-mail m.r.rose@kcl.ac.uk

Title; 82 characters

Abstract; 182 words

Text; 6,007 words

Additional material related to this article can be found on the *Neurology* Web site. Go to www.neurology.org and scroll down the Table of Contents for the _____issue to find the link for this article.

**ABSTRACT**

*Objectives:* To determine whether Gulf war veterans with neuromuscular symptoms that included weakness and fatigue had either i) objective correlates for muscle weakness or fatigue, or ii) any etiological explanation for such symptoms, and, if so iii) whether such objective measures or etiologic mechanisms were specific to Gulf war service.

*Subjects and methods:* Forty-nine ill Gulf war veterans with more than four neuromuscular symptoms (Gulf-ill) were compared with 26 Gulf-well veterans; 13 symptomatic Bosnian veterans (Bosnia-ill); and 22 symptomatic troops who were not deployed to the Gulf (Era-ill). Quantitative myometry was used to objectively measure weakness and fatigue. Subjects had an ischemic forearm exercise test, a sub-anerobic bicycle exercise test and a muscle biopsy.

*Results:* Quantitative strength and fatigue measures did not correlate with self perception of weakness or fatigue for any of our groups. No specific muscle biopsy abnormalities were found. There was no defect of adenylate deaminase or glycogenolysis found. Gulf-ill subjects did find the sub-anerobic bicycle exercise more effortful and generated significantly higher plasma lactate concentrations compared with Gulf-well subjects.

*Conclusion:* Since complaints of weakness and fatigue in unwell servicemen do not correlate with actual weakness or fatigue, explanations for these symptoms must lie outside of the neuromuscular system. Increased lactate production during sub-anerobic bicycle exercise reflects mitochondrial inefficiency, but it is unclear whether this reflects mitochondrial damage sustained during Gulf war service, or inactivity secondary to ill health.

**INTRODUCTION**

There has been considerable concern expressed by the media, politicians and ex-servicemen and women regarding possible long-term effects of service in the 1991 Persian Gulf War. Neuromuscular symptoms including weakness and fatigue have figured prominently in Gulf War symptom surveys(1-3). Despite this there have been no reports of weakness found on neurological examination of veterans and few reports of muscle biopsy results that might explain the basis for the symptoms of weakness or fatigue. One study on a self-reporting group of patients complaining of severe muscle fatigue, weakness or myalgias of six months duration or longer, sufficient to interfere with activities of daily living, assessed manual muscle strength of 36 muscle groups with MRC grading, as well as creatine kinase (CK) assays and open muscle biopsies. Muscle strength was normal. CK was raised in six subjects, five of whom were muscular black males who tend to have higher CK levels than the "normal".(4) There were three abnormal muscle biopsies; two with rare (<1%) tubular aggregates and a third with occasional necrotic fibres.(5) A neurophysiologic study reported by us did not show any abnormalities specific to Gulf War illness.(6) Neurophysiologic studies might not detect upper motor neurone weakness. However, it may be that the weakness complained of by veterans is relative to a greater degree of strength expected as normal for trained military personnel, and that a decline from this level while noticed by the individual still does not fall below that perceived as normal by the examining physician.

Fatigue is a subjective symptom with a complex phenomenology and the symptoms of fatigue and weakness may overlap, even though the two are physiologically different. Thus a fatigued subject may complain of weakness in the absence of actual weakness. Fatigue itself has a multitude of possible, sometimes overlapping, explanations including those due to neuromuscular pathology that would be expected to cause measurable decline in muscle strength with either sustained or repetitive muscle contraction i.e. fatigability. Recently it

has become possible to quantitate fatigability using the same equipment as used for fixed myometric assessment of muscle strength. Two such fatigue modalities can be measured. Static fatigue measures the rate of decline in the force generated by sustained muscle contraction over 30 seconds. Dynamic fatigue measures the drop in peak force generated by rhythmic contractions paced at one per second for 30 seconds. These measures allow quantitation of fatigue and should allow us to determine whether complaints of fatigue are related to objective abnormalities of muscle fatigue i.e. fatigability.

There are a number of possible theories as to the cause of fatigue in Gulf War solders. The use of acetylcholinesterase inhibitors as prophylaxis against organophosphate based chemical attack led to the suggestion that fatigue might result from neuromuscular junction dysfunction. However, our controlled neurophysiology study showed no evidence (by either repetitive stimulation or single fiber EMG) for neuromuscular junction dysfunction in either fatigued or non-fatigued Gulf War veterans.(6) But alternative explanations for fatigue remain, such as those due to induced defects in muscle glycolytic or mitochondrial metabolism. Mitochondrial dysfunction is an attractive potential explanation for fatigue in Gulf War related illness, since the multi-system nature of the symptoms induced by mitochondrial damage could explain other common symptoms including those related to CNS as well as neuromuscular function. Furthermore, those serving in the Gulf could have been exposed to a number of potential anti-oxidant chemicals that would have increased the likelihood of mitochondrial damage.

In this study we undertook to examine quantitative muscle strength and fatigue using fixed myometry and also assessed muscle biopsies. We wanted to answer the question as to whether self-reported or subjective complaints of weakness or fatigue were related to objective abnormalities of weakness or fatigue. We wanted to know if any of these

symptoms might relate to particular muscle biopsy abnormalities. We also wanted to determine whether complaints of fatigue could be explained by any abnormality of either muscle glycolytic or mitochondrial function. If any objective abnormalities were found, we also wanted to know whether this was specific to Gulf War service. We therefore used a randomly selected group of ill Gulf War veterans and compared them with both normal and ill military controls.

## METHODS

***Subjects.*** We aimed to randomly select 125 male UK military personnel (50 ill Gulf War veterans and three control groups of 25 veterans each) from our database of 8,195 symptomatic and asymptomatic service personnel from all branches of the armed services who had completed a detailed postal questionnaire, including the Medical Outcomes Study Short Form 36 (SF36)(7) and other health outcome questionnaires, as previously described(1). Although the SF-36 is a measure of quality of life it has very useful items of functioning. The Physical Function sub-scale consists of 10 items which asks the participant to rate his ability to carry out a range of common physical activities. We used it as a proxy marker for physical health. In the absence of any accepted case definition for Gulf War related illness we defined symptomatic servicemen as those who had an SF36 physical function sub-score of less than 72.2. This value of 72.2 represented the first decile of those service personnel who were in active service but not in an active theatre of conflict, and was the same as used to define disability in our previous studies(8). Moreover, in order to ensure that neuromuscular symptoms were represented we also had an additional requirement that they score more than four of the self-reported symptoms of neuromuscular dysfunction which included fatigue, joint stiffness, muscle weakness, myalgia at rest or after exercise, sensory symptoms, e.g. parasthesiae or numbness of fingers or toes, and autonomic symptoms, such as disturbances of bladder, bowel or sexual functions. These symptomatic subjects were derived from those who had served in the Gulf War (Gulf-ill group), and, as ill controls, those who had served in the Bosnia peacekeeping operation 1992 - 1997 (Bosnia-ill group) and those serving in the forces at the time, but not in either conflict (Era-ill group). The well controls were asymptomatic subjects who had served in the Gulf War with an SF36 physical function sub-score of more than 72.2 and scoring less than two of the self reported symptoms of neuromuscular dysfunction (Gulf-well group)(6).

***Clinical Assessment.*** As previously described all participants were examined over a two-day period by investigators who were not aware of their group status or clinical symptoms(6). The assessment included a more detailed survey of the participants' neuromuscular symptoms, including the presence or absence of fatigue, and its temporal relationship to exercise and the distribution and severity of their weakness, using a standard questionnaire filled in by the subjects. Subjects' responses to this questionnaire were categorised as indicating global, upper limb, lower limb, proximal or distal patterns of weakness. Subjects also completed the Chalder Fatigue Score and results were expressed as a total score and also as physical and mental fatigue sub-scores. Maximum voluntary isometric contraction (MVIC) was determined using a fixed myometry (QMA System, The Computer Source, Athens, Georgia, USA) by one of three clinical evaluators using methods that have demonstrated high test–retest reliability in facioscapulohumeral dystrophy.(9) Eight muscle groups were assessed bilaterally; shoulder abduction, elbow extension and flexion, hip flexion, ankle dorsiflexion, handgrip, knee flexion and knee extension in that order, so as to minimise positional changes and to reduce fatigue on muscle groups. The force measurements for each muscle group were transformed and standardized against the Gulf Well group allowing for age and height. This allowed expression of individual muscle MVIC scores as a z-score representing the number of standard deviations by which the strength of a particular muscle differed from that of the Gulf Well controls. These z scores were analysed individually, but were also combined to give composite regional MVIC scores as follows; global MVIC (all 16 muscle groups), upper limb MVIC (bilateral shoulder abduction, elbow flexion and extension, and hand grip), lower limb MVIC (bilateral hip flexion, knee flexion and extension, and ankle dorsiflexion), proximal MVIC (bilateral shoulder abduction and hip flexion), distal MVIC (bilateral hand grip and ankle dorsiflexion).

Hand-grip fatigue measures were performed using a Jaymar hand grip dynamometer while fatigue measures of other muscle groups used a fixed myometry system employing a force transducer (Super Mini Load Cell; Interface Inc, Scottsdale, AZ) with limb positions as described (9). Both were linked to QMA software (The Computer Source, Athens, Georgia, USA). Static fatigue was assessed during a single, 30-second sustained maximal contraction of the dominant ankle dorsiflexor, knee extensor, and handgrip. The static fatigue index for these muscle groups was calculated as;

Static Fatigue Index = $100\% \times \{1-(AUC_{5-30}/[F_{max}^{0-5} \times 25])\}$

(Where $AUC_{5-30}$ = Area Under the Curve between 5 and 30 seconds, $F_{max}^{0-5}$ = maximum force exerted between 0 to 5 seconds)

Dynamic fatigue was assessed in the non-dominant handgrip and ankle dorsiflexion during a series of rhythmic 1Hz maximal contractions, paced by a metronome, over 30 seconds. The dynamic fatigue index was calculated as;

Dynamic Fatigue Index = $100\% \times [1-(Fmax^{25-30}/Fmax^{0-5})]$

(Where $F_{max}^{25-30}$ = maximum force exerted between 25 to 30 seconds, $F_{max}^{0-5}$ = maximum force exerted between 0 to 5 seconds.)

Glycolytic function and myoadenylate deaminase activity were assessed using the ischemic forearm test (IFT).(10) Subjects performed isometric handgrip contractions using a Vigorimeter (Elmed Incorporated, Illinois, USA) at 80% of their maximum strength at a rate of 0.5hz while a blood pressure cuff was inflated around the upper arm at 20 mmHg above their systolic blood pressure. This test lasted until they were unable to reach 80% of their maximum grip despite encouragement. Venous blood samples were taken pre-exercise and at three, five and seven minutes post-exercise, stored on ice and immediately spun down at the conclusion of the test for assay of lactate and ammonia.

Mitochondrial function was assessed by the sub anaerobic threshold exercise test (SATET). Subjects exercised on an electronically braked bicycle (Tuntori, Nottingham, UK) for 15 minutes at 60 rpm at 90% of their predicted anaerobic threshold, as calculated from their weight.(11) Continuous cardiac monitoring was performed throughout and subjects were asked to rate perceived exertion using a Borg scale at 1 minute intervals (Gibson et al). Venous blood samples were taken pre-exercise, immediately post-exercise and 30 minutes after completion of exercise. Samples for lactate and ammonia were stored on ice and immediately spun down at the conclusion of the test. Samples for pyruvate were collected into perchloric acid (0.5 mL 8% perchloric acid + 1 mL blood), mixed thoroughly and centrifuged at 1500g for 10 minutes to deproteinise the sample. Pyruvate was measured using a direct enzymatic method (Sigma Diagnostics, Poole, UK) on the Cobas Mira analyser (Roche Diagnostics, Lewes, Sussex, UK). Plasma ammonia was measured using a direct enzymatic method (Trace ammonia method, Alpha Labs, Eastleigh, Hants, UK) on the Cobas Mira analyser (Roche Diagnostics, Lewes, Sussex, UK). Plasma lactate was measured using the YSI 2300 Stat glucose/lactate analyser (Clandon-YSI Ltd., Farnborough, Hants, UK). The IFT and SATET were performed on consecutive mornings with the subjects at rest and fasted for at least 12 hours prior to testing.

Muscle biopsies were performed under local anaesthetic using a muscle biopsy needle (Shuco International (London) Ltd) sampling from left vastas lateralis. Samples were frozen and processed for histology and histocytochemistry using the following stains; haematoxylin and eosin, Gomori trichrome, succinate dehydrogenase, ATPase at pH 9.6, 4.4 and 4.6, NADH, cytochrome oxidase, periodic acid Schiff, and Sudan Black. Quantitative morphometry was performed using the ATPase stained sections at pH 9.4; the lesser diameter of 100 randomly chosen fibers was measured with a graticule and each was

assigned as either a type 1 or type 2 fiber. Using this data atrophy and hypertrophy factors were calculated for each muscle biopsy.(12)

The study was approved by the institutional review board and participants consented for the study with a separate additional consent for the muscle biopsy.

***Statistical Analysis.***

Statistical analysis was performed in STATA (version 7.0). The distributions of the z-scores, both for individual muscles and for regional muscle groups, and the distributions of the static and dynamic fatigue indices of each of the muscle groups were examined across the four cohorts (Gulf-well, Gulf-ill, Era-ill, Bosnia-ill). When the normality and equal variances criteria were met, one-way ANOVA was carried out to assess mean differences between the four groups. Otherwise, non-parametric Kruskal-Wallis tests were used. Once a significant group effect was found, Bonferroni post-hoc comparisons were performed to assess which of the groups were different. We carried out chi-square tests to assess the association between cohort and complaints of weakness or fatigue. Chi-square tests were also used to examine the association between weakness and normality of muscle biopsy, whilst correlation analysis was carried out and the Pearson correlation coefficient was reported to assess the association between fatigue indices and Chalder global and physical fatigue sub-scores. ANCOVA was used to assess the main effect of weakness on average muscle strength, i.e. the difference between those who complained of weakness and those who did not, regardless of the group status, allowing for whether the individual was still in military service. In addition, the interaction between weakness and group was examined to assess whether differences between those with weakness and those without depended on the group status. Analysis of variance (ANOVA) was used to examine the differences in IFT and SATET

results between groups and the relationship between these results and complaints of fatigue and Chalder Fatigue Scores.

**RESULTS**

Of the 142 randomly selected service and ex-service personnel agreeing to attend, 110 consented to take part in this study and quantitative myometry was performed in 109, with 26 in the Gulf-well group, 49 in the Gulf-ill group, 12 in the Bosnia-ill group, and 22 in the Era-ill group. Forty-eight (44%) servicemen (18 Gulf-ill, 17 Gulf-well, 7 Era-ill and 6 Bosnia-ill) were still serving in the armed forces; the rest had left the services.

Table 1 gives the demographics of the 109 subjects and Table 2 summarises the numbers of subjects reporting weakness and the distribution of that weakness. Our recruitment strategy successfully obtained ill groups with significantly more complaints of weakness than reported in the well control group ($\chi^2 = 24.5$, df = 3, $p<0.001$). There was no significant difference in composite global MVIC scores between Gulf-ill, Bosnia-ill, Era-ill and Gulf-well ($F(3,97)=1.01, p=0.39$), (Figure 1). The same was true for individual muscle MVIC scores with the exception of non-dominant handgrip, which showed a statistically significant difference between Gulf-ill and Bosnia-ill group (mean difference = -1.3, 95%CI: (-2.4, -0.2), p = 0.04) with the Gulf-ill having the lowest score. To assess whether reduced MVIC scores were associated with reported patterns of weakness, we looked at differences in regional MVIC between those who complained of weakness and those who did not, regardless of their cohort status. We did this for five composite regional MVIC scores (as described above) and also for four specific muscle groups bilaterally (shoulder abduction, hip flexion, ankle dorsiflexion and handgrip) relating these to the appropriate reported distributions of weakness (Table 3). As there were significant differences in MVIC between those still in the military and those discharged, we considered this variable as a potential confounder. Allowing for military status, we found significant associations between MVIC score and reported weakness in three out of nine patterns of weakness namely global weakness (mean difference = -1.51, 95%CI: (-2.69, -0.34), p = 0.012), proximal weakness (mean difference = -0.32, 95%CI: (-0.64, -0.007), p = 0.04) and shoulder abduction (mean

difference = -0.35, 95%CI: (-0.64, -0.06), p = 0.02). For each of these three significant associations there was no significant interaction between the weakness and cohort status; global weakness ($F(1, 93) = 1.59$, $p = 0.21$), proximal weakness ($F(2, 92) = 0.09$, $p = 0.91$), proximal UL weakness ($F(3, 91) = 0.42$, $p = 0.74$). This implies that the above differences in muscle strength were the same across the four cohorts, i.e. the association between reduced muscle strength and a particular distribution of weakness is not group specific. We did not find any significant evidence for association between regional MVIC score and the remaining six distributions of weakness as shown in Table 3.

All three ill groups had a significantly higher Chalder Fatigue Score, whether total score or physical or mental sub-scores, than the well group regardless of deployment status (Table 4). No association was found between cohort status and static or dynamic fatigue measures for each individual muscle (Table 5). Table 6 shows that there were no significant associations between the static fatigue score for any individual muscle and the Chalder Fatigue Score, with the exception of a small significant association between static fatigue for handgrip and the Chalder Fatigue Score total score and physical fatigue sub-score. There was a similar lack of association between dynamic fatigue measures and Chalder Fatigue Score (Table 6).

All subjects showed the expected rise in lactate and ammonia concentrations during the ischemic forearm test (IFT) with no difference between the groups (lactate; $F = 0.66$, $df = 3$, $p = 0.58$, ammonia; $F = 0.25$, $df = 12$, $p = 0.99$).

All subjects showed similar baseline lactate and pyruvate concentrations during the sub anaerobic threshold exercise test (SATET), but there were differences in lactate concentrations between groups at 15 and 30 minutes (Figure 2). Lactate concentration at 15 minutes was significantly higher for Gulf-ill (mean diff. = 1.08, 95%CI: 0.70 - 1.46, $p < 0.001$) and Era-ill groups (mean diff. = 0.78, 95%CI: 0.08 – 1.49, $p = 0.03$) than that of Gulf-

well group, while those in Bosnia-ill tended to have higher lactate concentration than those in Gulf-well group (mean diff. = 0.40, 95%CI: (-0.07, 0.87), p = 0.09). Post-exercise lactate concentration at 30 minutes was higher for Gulf-ill than that of Gulf-well (mean diff. = 0.23, 95%CI: 0.07 - 0.38, p = 0.004), but Era-ill (mean diff. = 0.15, 95%CI: (-0.10, 0.41), p = 0.23) and Bosnia-ill (mean diff. = 0.02, 95%CI: (-0.24, 0.28), p = 0.86) did not differ significantly from the Gulf-well group. Pyruvate concentrations showed no significant time by group interaction (F = 1.11, df = 6, p = 0.36), suggesting that there was no difference between the groups in terms of baseline, peak or recovery pyruvate concentrations.

Mean peak Borg scores for each cohort are given in Table 7. Peak Borg scores registered during the bicycle exercise test were significantly higher for the Gulf-ill compared with the Gulf-well cohort (mean difference = 1.0, 95%CI: 0.50 – 1.51; p = 0.01) suggesting that Gulf-ill subjects found this exercise more of an effort. The peak Borg score was only weakly associated with the Chalder Fatigue Score (Spearman's rho = 0.27, (p = 0.004)) and this was mainly due to the association with the physical fatigue sub-score (Spearman's rho = 0.34, (p=0.0003)). There was a significant, but moderate, association between peak Borg score and lactate concentration for 15 and 30 minutes (Spearman's rho = 0.31, p = 0.001). Peak Borg scores did not generally relate to either static or dynamic fatigue myometry scores as only a weak negative association was found with static fatigue for ankle dorsiflexion (Spearman's rho = -0.20, p = 0.06) and dynamic fatigue for hand grip (Spearman's rho = -0.20, p = 0.05).

Thirty seven subjects consented to a muscle biopsy with no significant differences in the willingness to consent between the four groups ($\chi^2 = 2.13$, df = 3, p = 0.54). Three biopsies were technically inadequate (but were classed as abnormal for the purposes of statistical analysis), 18 were normal and 16 showed a variety of abnormalities (Table 8). There was no

significant association between those muscle biopsies showing any kind of abnormality with any reported weakness ($\chi^2$ = 0.01, df = 1, p = 0.92) and none of the abnormalities were associated with any particular cohort. Suggestions of mitochondrial abnormality arose in two biopsies; one biopsy from a Era-ill subject showed three cytochrome oxidase negative fibers, while another biopsy from a Gulf-ill subject showed an increase in trichrome staining not amounting to "ragged red change". Neither of these subjects complained of weakness and their MVIC scores were not reduced.

There was no association between fiber type ratio and cohort (p = 0.89), nor any association between fiber diameters and cohort (p = 0.28 for type 1 and p = 0.16 for type 2). Atrophy factors for type 1 and type 2 fibers were not related to cohort (p = 0.90 for type 1 and p = 0.35 for type 2) and similarly hypertrophy factors for type 1 and type 2 fibers were not related to cohort (p = 0.62 for type 1 and p = 0.85 for type 2). Irrespective of cohort, atrophy factors were not related to illness, either for type 1 fibers (median atrophy factor for the ill group: 37.1, range: 0 – 290.9; median for the well group: 62.5, range: 0 – 151.5, p = 0.72) or for type 2 fibers (median atrophy factor for the ill group: 47.6, range: 0 – 425.5, median for the well group: 106.4, range: 0 – 307.7, p = 0.35). Finally, there was no association between type 1 or 2 fiber hypertrophy factors and quantitative strength scores.

**DISCUSSION**

The purpose of the study was to identify any possible neuromuscular causes for ill health in Gulf War veterans. There is no accepted case definition for Gulf War ill health, so we pragmatically developed a definition that selected veterans who had disability and prominent neuromuscular symptoms, reasoning that if neuromuscular causes exist for Gulf related illness this group would be most likely to have neuromuscular abnormalities. The random selection of servicemen with poor physical functioning and neuromuscular symptoms from a large database of Gulf war veterans together with closely matched symptomatic and asymptomatic controls give us the advantage of being able to determine whether symptoms relate to objective abnormalities and whether any abnormal objective measures are specific to Gulf war service. Our case definition successfully produced ill cohorts that had significantly higher levels of self reported complaints of weakness and lower Chalder Fatigue Scores compared with the well control group.

Our original hypothesis was that any self-perceived weakness might be within the normal range clinically, and hence not be detected as weakness in traditional manual muscle testing, but that such weakness might be detectable by quantitative myometry especially with reference to matched military controls. We did find that those men still in service were stronger than those discharged from military service thus supporting our use of military controls. The study revealed a number of statistically significant associations between reported weakness and reduced strength as measured by quantitative myometry. However, the question is whether these statistically significant associations are clinically significant. Regardless of this judgement what is clearly demonstrated is that any such associations, whether statistically or clinically significant, do not relate specifically to Gulf War service. Thus we are discussing whether or not there is a general “post-war syndrome” and not a

specific Gulf war effect. This would be consistent with the finding that a post-war syndrome has been a feature of several previous military conflicts.(13)

None of the groups had statistically significantly different individual muscle or composite MVIC scores with the exception of non-dominant handgrip being 1.3 standard deviations lower in Gulf-ill group as compared with the Bosnia-ill group. This difference is less than the traditionally accepted clinically significant difference of greater than 2.0 standard deviations and is not explained by the presence of either median or ulnar neuropathy, which was only clinically evident in three Gulf-ill subjects (two with median and one with ulnar nerve compression). We increased the sensitivity of the study in picking up reduced MVIC scores in those with reported weakness by looking at regional and specific muscle MVIC scores comparable with the distribution of reported weakness. We found statistically significant differences in such MVIC scores for four distributions of weakness. For three of these patterns of weakness the difference was less than one standard deviation (-.32, -0.35 and –0.87). Two subjects, one Gulf-ill and one Era-ill, complained of global weakness and had global MVIC scores that were a mean of –1.51 (–0.34 and –2.69) standard deviations below that obtained by subjects with no reported weakness. One needs to be cautious in ascribing true statistical significance to the mean difference between a group containing just 2 individuals and a larger group of 98 individuals (those that did not have any weakness). We therefore feel that none of our data reflects clinically significant objective weakness to correlate with self-perceived weakness.

More recently it has been suggested that amyotrophic lateral sclerosis (ALS), for which the cardinal symptom is weakness, may be more likely in Gulf war service personnel.(14), (15), (16) While in ALS it is objective rather than subjective weakness that makes the diagnosis, it is recognised that weakness due to ALS may only be detectable using quantitative myometry before it is detectable on manual muscle testing. Given the rarity of ALS it is not

surprising that we found no cases of ALS in any of our subjects, but complaints of weakness are much more common than is ALS and it is reassuring to realise that, in all probability, the majority of such complaints are not associated with abnormal quantitative myometry that would suggest pre-clinical ALS.

Both static fatigue and dynamic fatigue muscle measures were unrelated to the ill status of subjects. Neither fatigue measure correlated with Chalder Fatigue Scores, with the exception of a small association between static fatigue for hand grip and the total Chalder Fatigue Score . The Chalder Fatigue total score includes elements of mental and physical fatigue and it would seem logical that measures of motor fatigue might better correlate with the Chalder Fatigue physical sub-score. However, with the exception of a small association with static fatigue for hand grip, the Chalder Fatigue physical sub-scores did not correlate with the motor fatigue measurements. Thus it seems that individuals have a measurable perception of fatigue, whether mental or physical, that does not relate to any objective measure of muscle fatigue.

Our knowledge of the relationship between different fatigue states and the quantitative motor fatigue measures is limited, as the latter have only recently become available. A study looking at quantitative strength and fatigue measurements in patients with multiple sclerosis showed that quantifiable fatigue was seen in these subjects, and that it sometimes occurred in strong muscles. However, this study did not relate quantitative fatigue to any self rated scale of fatigue.(17) Some models of fatigue suggest that muscle fatigue may have a central and a peripheral element. The peripheral element may relate to defects in neuromuscular transmission or muscle function, such as biochemical deficiency, while the central element is less well understood, but may result in a reduced upper motor neuron drive to the lower motor neuron units. It has been suggested that central fatigue might be distinguished from peripheral fatigue by the ability to obtain additional muscle strength over and above apparent

maximal voluntary isometric contraction using central stimulation, for example, by magnetic coils. Such a model clearly does not apply in our subjects where both quantitative muscle strength (maximal voluntary isometric contraction) and quantitative fatigue measures are not reduced.

This study does not address the issue of mental as opposed to physical fatigue, which might account for the low mental fatigue sub-scores of the Chalder Fatigue Scale. It is possible that mental fatigue, while not contributing to any centrally derived muscle fatigue, nevertheless induces a subjective feeling of physical fatigue. It has been suggested that the perception of fatigue in the absence of true weakness or fatigability, might be explicable on the basis of a primary disturbance of the sense of effort, and if so one might predict that peak Borg scores during exercise should correlate with Chalder Fatigue Scores.(18) In fact our study showed only a weak association between peak Borg scores in exercise, and Chalder Fatigue scores, this being mainly with the physical fatigue sub-score.

With respect to the potential causes for peripheral muscle fatigue we have already shown that there is no evidence for any neuromuscular transmission defect in “ill” subjects regardless of cohort.(6) This study showed no evidence for any defect in muscle glycolytic metabolism or myoadenylate deaminase activity as judged by the normal lactate and ammonia concentration responses during the ischemic forearm tests. However, the sub anaerobic threshold exercise test (SATET) showed a significantly higher peak and recovery lactate concentration in Gulf-ill subjects as compared with Gulf-well and Bosnia-ill subjects. This suggests that mitochondrial respiratory chain function is relatively impaired in Gulf-ill subjects. In mitochondrial disease, acquired or genetic defects of the mitochondrial oxidative phosphorylation system can lead to multi-system disorders including muscle

weakness and fatiguability. In many, but not all cases, the defect of oxidative phosphorylation triggers an unsuccessful attempt at compensation by way of mitochondrial proliferation and this may be seen in muscle biopsy samples as ragged red fibers. The elevated lactate concentrations seen in the Gulf-ill group are not of the magnitude seen in traditional mitochondrial muscle disease associated with ragged red fibers in muscle biopsies. None of our subjects had ragged red fibers on their muscle biopsies, but subtle mitochondrial abnormalities were seen in two cases. It may be that the abnormal SATET results reflect a subtle mitochondrial abnormality that is not sufficient to trigger the full picture of a mitochondrial disease in muscle, but that is sufficient to cause some of the symptoms described in the Gulf-ill subjects. Having said that, the fact is that we have not been able to find objective evidence for weakness or fatigue that would be expected for a mitochondrial abnormality. This study does not tell us whether any such mitochondrial abnormalities are more or less pronounced in the central nervous system, nor whether these would account for any CNS derived symptoms.

A second explanation for the impaired SATET results may be that these reflect the "unfitness" or "de-trained state" of the subjects caused by their symptoms. Untrained or unfit subjects do have less efficient mitochondrial respiratory chain function as compared with fit or active subjects(19). In other words the mitochondrial inefficiency may be the refection of, rather than the cause of, the ill state of the subjects. In support of the suggestion that Gulf-ill subjects were less fit is the fact that they were significantly heavier than the control groups, perhaps as a result of a more sedentary life style induced by their symptoms, or having left the armed forces. Gulf-ill subjects also had a significantly higher perception of the effort required to complete the bicycle exercise test with a general, but not cohort specific, association between peak Borg scores and the extent of the rise in lactate concentrations. This again may reflect the unfit state.

One would not expect physiologically induced mitochondrial inefficiency, relating to lack of fitness, to result in objective physical weakness or fatigue, nor would it trigger ragged red fibers in the muscle biopsy. If this mitochondrial inefficiency was merely a reflection of the "ill" state and the unfitness induced by this, then one might suppose that the same effect would be seen with all three ill groups. In fact only the Gulf-ill group showed a significant difference in the SATET compared with the Gulf-well group. However, while there was no significant difference between the Gulf-ill and Era-ill groups and looking at the graphs (Figure 1) one can see that the ill cohorts all showed a trend towards higher peak and recovery lactate concentrations compared with the Gulf-well cohort. Thus it may be that the significant result seen with the Gulf-ill group reflects the degree to which they were unfit and that this in turn may relate to the number of symptoms that they have. Bearing in mind that our ill cohorts did not have objective impairment of muscle function it is possible that perceived muscle impairment or other non-muscle symptoms (e.g. central nervous system or non-neurological symptoms) were contributing to this unfitness.

If mitochondrial abnormality was indeed the cause rather than the effect of symptoms how might this have come about? Gulf service personnel were exposed to a number of factors that might damage mitochondrial DNA (mtDNA) given its increased susceptibility to mutational damage when exposed to anti-oxidants. Such damage would be expected to be random in its distribution within the mtDNA genome resulting in multiple deletions and mutations. MtDNA mutations become fixed in post-mitotic tissue, such as muscle and brain, where there is less opportunity for mtDNA mutations to be selected out during successive cell divisions. One might therefore predict that the main effects of such mitochondrial damage would be seen in muscle and in the central nervous system which does seems to equate with many of the symptoms described in Gulf-well veterans. With age one tends to

accumulate mtDNA mutations and so mitochondrial function declines as one get older. Thus one would predict that the effects of any pre-existing war related mitochondrial damage would worsen with time and result in increasing symptoms.

We found muscle biopsy abnormalities in 16 out of 34 samples, most of which in themselves would be considered minor and of doubtful, or uncertain, significance in the normal clinical setting. But regardless of any subjective view as to the clinical significance of these biopsy abnormalities, the design of our study allows us to say that these abnormalities are not related to complaints of weakness, lower MVIC scores, or military service, whether in the Gulf war or not. It has been suggested that mitochondrial abnormalities may have resulted from toxic exposure during Gulf War service, but histochemical abnormalities suggestive of mitochondrial abnormality were only found in two subjects, both from ill groups, neither of whom complained of weakness, and with only one having served in the Gulf War. This Gulf War subject had a mild increase in peripheral trichrome staining which is a subjective finding of arguable significance in terms of mitochondrial function. However, the three cytochrome oxidase negative fibers seen in the Era-ill subject would be regarded as definitely abnormal in a 46 year old. Quantitative morphometric analysis did not show fiber type ratio or fiber diameters to be associated with cohort. Muscle fiber diameters tend to reflect muscle activity with atrophy being associated with disuse and hypertrophy being associated with increased activity, especially for type 2 fibers. However, neither atrophy nor hypertrophy factors were associated with illness, nor were they associated with perceived or actual weakness.

Thus quantitative assessments and the muscle biopsies did not show abnormalities that related to the self-perception of weakness or fatigue, and we have not found a Gulf War specific effect. This finding is supported by the lack of any specific neurophysiologic

abnormalities in our subjects.(6) It also confirms the findings of the only similar, but uncontrolled, study of neuromuscular symptoms in Gulf war veterans.(5) This study adds to the information given by the neurophysiology study in that quantitative muscle strength scores were they abnormal, could have reflected upper motor neuron weakness that might not have been detected by neurophysiology. The lack of any specific Gulf War effect is in keeping with the fact that although the prevalence of symptoms in Gulf War veterans is higher as compared with veterans of the Bosnia campaign and serving, but not deployed troops, the pattern of symptoms is the same for all three groups.(1) The failure to correlate symptoms of weakness or fatigue with objective assessments of abnormalities of weakness or muscle biopsy abnormalities should not be taken as evidence that such self-reported symptoms are not genuine nor worthy of concern. However, it does suggest that their origin does not lie in the neuromuscular system. We have found that Gulf-ill subjects found exercise more effortful and that it generated more lactic acid than with controls, but we are unable to say conclusively whether this was the result of lack of fitness or due to subtle, perhaps environmentally induced, mitochondrial damage. Ultimately the discussion as to whether apparent mitochondrial inefficiency is cause or effect could be resolved by direct analysis of mtDNA using real-time polymerase chain reaction techniques to see whether the amount of mtDNA mutations in post-mitotic tissue e.g. muscle, is greater in ill service personnel as compared with matched controls.

**Acknowledgement**

This study was funded by the UK Ministry of Defence (MOD). The MOD have had no input into the design, conduct, analysis or reporting of the study. We would like to thank all the veterans who gave so freely of their time.

**TABLES**

Table 1. Clinical characteristics of the study cohort at the time of original selection. Values represent means ± standard deviation

| Clinical Group | Total number | Age (years) | Height (cm) | Weight (kg) |
|---|---|---|---|---|
| GI | 49 | 37.5 ± 6.9 | 177 ± 8.8 | 91.8 ± 13.4[a] |
| GW | 26 | 34.7 ± 5.5 | 176 ± 5.8 | 80.3 ± 14.5 |
| BI | 12 | 32.8 ± 6.8 | 175 ± 6.0 | 77.7 ± 14.7 |
| EI | 22 | 37.5 ± 6.6 | 176 ± 6.6 | 86.2 ± 14.2 |

[a] $p < 0.01$ compared to mean body weight of the Gulf-well and Bosnia veterans

GW = Gulf-well, GI = Gulf-ill, EI = Era-ill, BI = Bosnia-ill

Table 2. Distribution of weakness reported by participants

| | **No weakness** | **Global Weakness** | **UL Weakness** | **LL Weakness** | **Proximal weakness** | **Distal weakness** | **UL proximal weakness** | **UL distal Weakness** | **LL proximal Weakness** | **LL distal Weakness** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Total(N = 109)** | 44(40%) | 2(3%) | 19(29%) | 13(20%) | 25(38%) | 4(6%) | 32(49%) | 27(41%) | 54(83%) | 16(25%) |
| **GI(N = 49)** | 16(33%) | 1(50%) | 14(74%) | 8(61%) | 17(68%) | 2(50%) | 19(60%) | 18(67%) | 30(55%) | 8(50%) |
| **GW(N = 26)** | 21(81%) | 0 | 0 | 0 | 0 | 0 | 2(6%) | 1(4%) | 2(4%) | 1(6%) |
| **BI(N = 12)** | 3(25%) | 0 | 2(10%) | 1(8%) | 3(12%) | 0 | 4(12%) | 2(7%) | 7(13%) | 2(13%) |
| **EI( N = 22)** | 4(18%) | 1(50%) | 3(16%) | 4(31%) | 5(20%) | 2(50%) | 7(22%) | 6(22%) | 15(28%) | 5(31%) |

Note: Percentages refer to the percent of the cohort with given pattern of weakness (some subjects had more than one pattern of weakness).

GW = Gulf-well, GI = Gulf-ill, EI = Era-ill, BI = Bosnia-ill

Table 3; Mean Differences in Regional MVIC scores related to complaint of weakness (bold indicates statistically significant association)

| **MVIC score Mean difference 95%CI** | **Global Weakness N = 2** | **UL Weakness N = 19** | **LL global Weakness N = 13** | **Proximal weakness N = 25** | **Distal weakness N = 4** | **UL proximal weakness N = 32** | **UL distal Weakness N = 27** | **LL proximal Weakness N = 54** | **LL distal Weakness N = 16** |
|---|---|---|---|---|---|---|---|---|---|
| **Global MVIC** | **-1.51* (-2.69, -0.34)** | | | | | | | | |
| **UL MVIC** | | -0.19 (-0.57, 0.18) | | | | | | | |
| **LL MVIC** | | | -0.27 (-0.88, 0.33) | | | | | | |
| **Proximal MVIC** | | | | **-0.32* (-0.64, -0.007)** | | | | | |
| **Distal MVIC** | | | | | -0.87 (-1.76, 0.02) | | | | |
| **UL proximal MVIC ie shoulder abduction** | | | | | | **-0.35* (-0.64, -0.06)** | | | |
| **UL Distal MVIC ie handgrip** | | | | | | | -0.27 (-0.84, 0.29) | | |
| **LL Proximal MVIC ie hip flexion** | | | | | | | | -0.28 (-0.73, 0.16) | |
| **LL Distal MVIC ie ankle dorsiflexion** | | | | | | | | | -0.48 (-0.97, 0.009) |

Table 4; Chalder Fatigue Scores by cohort

| | **Chalder Fatigue Score** mean, (SD) | | | | |
|---|---|---|---|---|---|
| | **GW** | **GI** | **EI** | **BI** | **Statistic*** |
| **Total score** | 11.8(6.3) | 15.6(9.6) | 12.7(11.2) | 19.4(6.1) | $\chi^2 = 9.3$, p = 0.025 |
| physical sub-score | 6.1(3.3) | 8.3(5.3) | 7.5(6.8) | 10.3(3.9) | $\chi^2 = 6.5$, p = 0.08 |
| mental sub-score | 4.0(2.4) | 5.1(3.5) | 3.5(3.1) | 6.4(2.7) | $\chi^2 = 7.2$, p = 0.065 |

* Based on Kruskal-Wallis non-parametric test

GW = Gulf-well, GI = Gulf-ill, EI = Era-ill, BI = Bosnia-ill

Table 5. Static fatigue and dynamic fatigue by cohort

| **Static fatigue**, mean % decline, (SD) | | | | | |
|---|---|---|---|---|---|
| **Muscle** | **GW** | **GI** | **EI** | **BI** | **statistic** |
| **Handgrip** N = 98 | 35.2(12.3) N = 26 | 39.2(14.0) N = 44 | 37.9(9.9) N = 20 | 30.7(7.5) N = 8 | F = 1.33, df = 3, p = 0.27 |
| **Knee extension** N = 87 | 10.1(17.1) N = 26 | 17.1(20.9) N = 36 | 10.7(10.8) N = 17 | 13.7(9.5) N = 8 | F = 0.99, df = 3, p = 0.40 |
| **Ankle dorsiflexion** N = 96 | 19.8(8.0) N = 26 | 17.7(14.8) N = 42 | 14.8(9.9) N = 19 | 17.6(7.1) N = 9 | F = 0.66, df = 3, p = 0.58 |
| **Dynamic fatigue**, mean % decline, (SD) | | | | | |
| **Muscle** | **GW** | **GI** | **EI** | **BI** | **statistic** |
| **Handgrip** N = 98 | 23.6(12.5) N = 25 | 24.3(14.5) N = 43 | 26.5(10.1) N = 21 | 17.5(9.8) N = 9 | F = 1.05, df = 3, p = 0.37 |
| **Ankle dorsiflexion** N = 94 | 14.5(12.1) N = 26 | 10.8(19.8) N = 40 | 16.6(12.3) N = 19 | 14.7(6.04) N = 9 | F = 0.71, df = 3, p = 0.55 |

GW = Gulf-well, GI = Gulf-ill, EI = Era-ill, BI = Bosnia-ill

Table 6. Relationship between static and dynamic fatigue for each muscle and Chalder Fatigue Score

| **Static fatigue** (Pearson's correlation coefficient) | | | | |
|---|---|---|---|---|
| **Chalder Fatigue Scores** | | **Handgrip** | **Knee extension** | **Ankle dorsiflexion** |
| | Total score | 0.23 (S) n = 98 | 0.09 (NS) n = 87 | 0.02 (NS) n = 96 |
| | physical sub-score | 0.25 (S) n = 98 | 0.10 (NS) n = 87 | 0.007 (NS) n = 96 |
| | mental sub-score | 0.16 (NS) n = 98 | 0.05 (NS) n = 87 | 0.03 (NS) n = 96 |
| **Dynamic fatigue** (Pearson's correlation coefficient) | | | | |
| | Total score | -0.16 (NS) n = 98 | ND | -0.14 (NS) n = 94 |
| | physical sub-score | -0.18 (NS) n = 98 | ND | -0.13 (NS) n = 94 |
| | mental sub-score | -0.13 (NS) n = 98 | ND | -0.17 (NS) n = 94 |

NS: Non-significant
S: Significant at 5% level
ND: not done

Table 7, Comparison of peak Borg score, during bicycle exercise, by cohort

| | **GW(N=26)** | **GI(N=47)** | **EI(N=20)** | **BI(N=12)** | **Statistic** |
|---|---|---|---|---|---|
| Peak Borg Mean(SD) | 2.69(0.78) | 3.70(1.38) | 3.55(1.64) | 2.75(1.14) | F=4.41,df= 3,p = 0.006 |

GW = Gulf-well, GI = Gulf-ill, EI = Era-ill, BI = Bosnia-ill

Table 8; Muscle biopsy results

| | Gulf well | Gulf ill | Bosnia ill | Era ill | Total |
|---|---|---|---|---|---|
| Normal | 3 | 8 | 5 | 2 | 18 |
| Increase type 1 fibers | 3 | 4 | 1 | 0 | 8 |
| Atrophic fibers | 1 | 3 | 1 | 2 | 7 |
| Fiber size variability | 0 | 2 | 1 | 0 | 3 |
| Hypertrophic fibers | 0 | 2 | 0 | 0 | 2 |
| Internal nuclei | 0 | 0 | 0 | 2 | 2 |
| Endomysial fibrosis | 0 | 1 | 0 | 1 | 2 |
| Type 2 fiber atrophy | 0 | 0 | 1 | 0 | 1 |
| Fiber splitting | 0 | 1 | 0 | 0 | 1 |
| Increased trichrome staining | 0 | 1 | 0 | 0 | 1 |
| Moth-eaten NADH | 0 | 1 | 0 | 0 | 1 |
| Intrafasicular fibrosis | 0 | 1 | 0 | 0 | 1 |
| COX negative fibers | 0 | 0 | 0 | 1 | 1 |

**FIGURES**

Figure 1. Standardised z-scores for composite MVIC scores by cohort

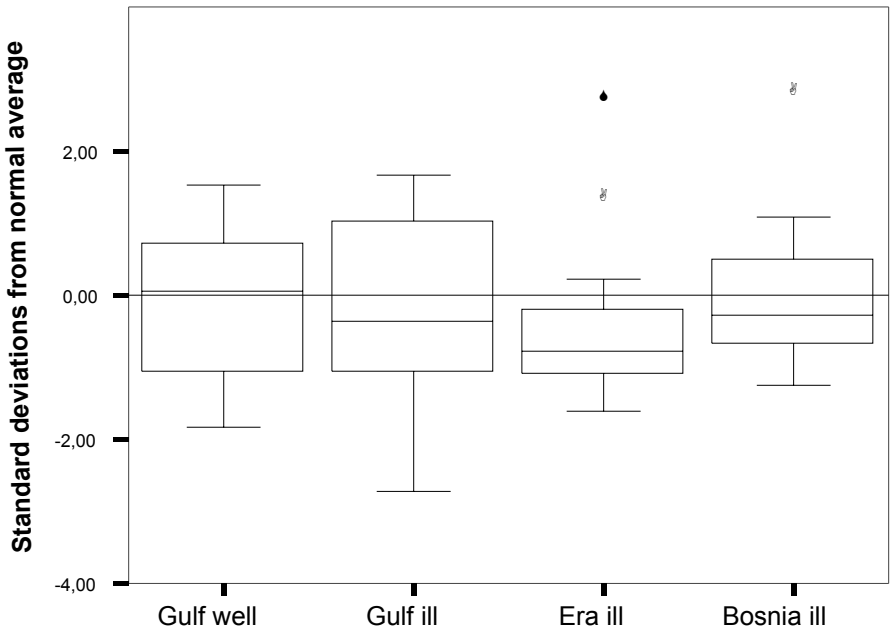
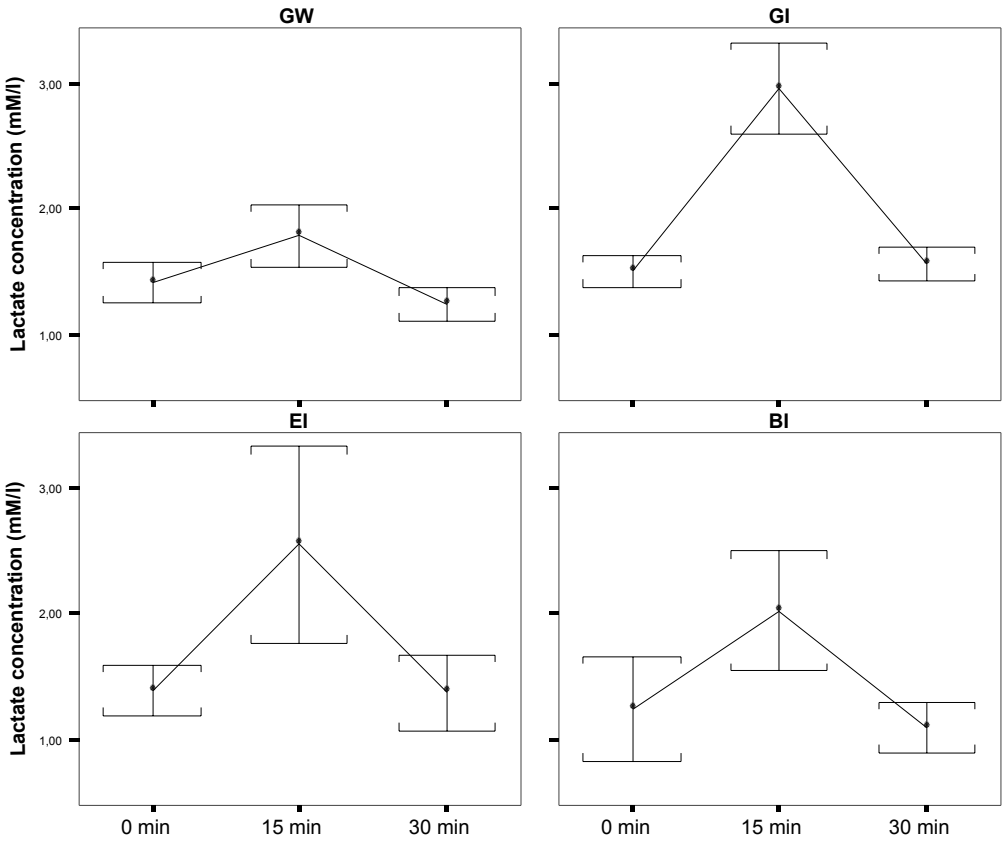


Figure 2; Bicycle exercise test. Lactate concentration by group over time




GW = Gulf-well, GI = Gulf-ill, EI = Era-ill, BI = Bosnia-ill

Reference List

(1) Unwin C, Blatchley N, Coker W, et al. Health of UK servicemen who served in Persian Gulf War. *Lancet* 1999 Jan 16;353(9148):169-178.

(2) Kang HK, Mahan CM, Lee KY, Magee CA, Murphy FM. Illnesses among United States veterans of the Gulf War: a population-based survey of 30,000 veterans. *J Occup Environ Med* 2000 May;42(5):491-501.

(3) Cherry N, Creed F, Silman A, et al. Health and exposures of United Kingdom Gulf war veterans. Part I: The pattern and extent of ill health. *Occup Environ Med* 2001 May;58(5):291-298.

(4) Sherwood RA, Lambert A, Newham DJ, Wassif WS, Peters TJ. The effect of eccentric exercise on serum creatine kinase activity in different ethnic groups. *Ann Clin Biochem* 1996 Jul;33 ( Pt 4):324-329.

(5) Amato AA, McVey A, Cha C, et al. Evaluation of neuromuscular symptoms in veterans of the Persian Gulf war. *Neurology* 1997;48:4-12.

(6) Sharief MK, Pridden J, Delamont RS, et al. Neurophysiologic analysis of neuromuscular symptoms in UK Gulf War veterans. A controlled study . *Neurology* 2002;59:1518-1525.

(7) Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992 Jun;30(6):473-483.

(8) Ismail K, Kent K, Brugha T, et al. The mental health of UK Gulf war veterans: phase 2 of a two phase cohort study. *BMJ* 2002 Sep 14;325(7364):576.

(9) Personius KE, Pandya S, King WM, Tawil R, McDermott MP. Facioscapulohumeral dystrophy natural history study: standardization of testing procedures and reliability of measurements. The FSH DY Group. *Physical Therapy* 1994;74:253-263.

(10) Sinkeler SP, Wevers RA, Joosten EM, et al. Improvement of screening in exertional myalgia with a standardized ischemic forearm test. *Muscle Nerve* 1986 Oct;9(8):731-737.

(11) Nashef L, Lane RJ. Screening for mitochondrial cytopathies: the sub-anaerobic threshold exercise test (SATET). *Journal of Neurology, Neurosurgery & Psychiatry* 1989 Sep;52:1090-1094.

(12) Brooke MH, Engel WK. The histographic analysis of human muscle biopsies with regard to fiber types. 1. Adult male and female. *Neurology* 1969 Mar;19(3):221-233.

(13) Jones E, Hodgins-Vermaas R, McCartney H, et al. Post-combat syndromes from the Boer war to the Gulf war: a cluster analysis of their nature and attribution. *BMJ* 2002 Feb 9;324(7333):321-324.

(14) Haley RW. Excess incidence of ALS in young Gulf War veterans. *Neurology* 2003 Sep 23;61:750-756.

(15) Horner RD, Kamins KG, Feussner JR, et al. Occurrence of amyotrophic lateral sclerosis among Gulf War veterans. *Neurology* 2003 Sep 23;61:742-749.

(16) Rose MR. Gulf War service is an uncertain trigger for ALS. *Neurology* 2003 Sep 23;61:730-731.

(17) Schwid SR, Thornton CA, Pandya S, et al. Quantitative assessment of motor fatigue and strength in MS. *Neurology* 1999 Sep 11;53(4):743-750.

(18) Lawrie SM, MacHale SM, Power MJ, Goodwin GM. Is the chronic fatigue syndrome best understood as a primary disturbance of the sense of effort? *Psychol Med* 1997 Sep;27(5):995-999.

(19) Coyle EF, Martin WH, III, Bloomfield SA, Lowry OH, Holloszy JO. Effects of detraining on responses to submaximal exercise. *J Appl Physiol* 1985 Sep;59(3):853-859.