Psychological Medicine, 2006, **36**, 735–747. © 2006 Cambridge University Press
doi:10.1017/S0033291705006975 First published online 26 January 2006 Printed in the United Kingdom

REVIEW ARTICLE

Systematic review of multi-symptom conditions in Gulf War veterans

HOLLIE V. THOMAS[1], NICOLA J. STIMPSON[1], ALISON L. WEIGHTMAN[2], FRANK DUNSTAN[3] AND GLYN LEWIS[4]*

[1] *Department of Psychological Medicine, School of Medicine, Cardiff University, UK;* [2] *Department of Information Services, School of Medicine, Cardiff University, UK;* [3] *Department of Epidemiology, Statistics and Public Health, School of Medicine, Cardiff University, UK;* [4] *Academic Unit of Psychiatry, University of Bristol, UK*

ABSTRACT

Background. Gulf War veterans have a number of health complaints. We therefore decided to carry out a systematic review to identify and summarize the findings from studies that have assessed multi-symptom conditions in Gulf War veterans and in an unexposed comparison group.

Method. Studies published between January 1990 and May 2004 were identified by searching a large number of electronic databases. Reference lists and websites were also searched and key researchers were contacted. Studies were included if they compared the prevalence of chronic fatigue syndrome, multiple chemical sensitivity, CDC-defined chronic multi-symptom illness, fibromyalgia, or symptoms of either fatigue or numbness and tingling in Gulf War veterans and non-Gulf veterans. A total of 2401 abstracts were independently reviewed by two authors.

Results. Twenty-three publications fulfilled the inclusion criteria. Gulf deployment was most strongly associated with chronic fatigue syndrome (OR 3·8, 95% CI 2·2–6·7). Gulf War veterans were also approximately three and a half times more likely than non-Gulf veterans to report multiple chemical sensitivity or chronic multi-symptom illness as defined by CDC. The methodological quality of the studies varied but the later and larger studies were of a high methodological standard with robust sampling strategies, adequate response rates and good adjustment for confounders.

Conclusions. The results support the hypothesis that deployment to the Gulf War is associated with greater reporting of multi-symptom conditions.

INTRODUCTION

Ever since the end of the Gulf War in 1991, its veterans have expressed concern over a number of health complaints. Symptoms that have been reported by Gulf War veterans include sleep problems, headache, lack of concentration, muscle stiffness and stomach upsets. This has led to the investigation of whether a specific 'Gulf War Syndrome' exists (Haley *et al.* 1997; Ismail *et al.* 1999). These are common symptoms that often lead to presentation in primary care and hospital settings (Kroenke & Price, 1993). As a result, many medical specialities have developed and defined various multi-symptom syndromes that are medically unexplained. These include chronic fatigue syndrome, fibromyalgia and multiple chemical sensitivity (Wessely *et al.* 1999). There is still controversy over the nosological status of these syndromes, but the existence of agreed case-definitions and the associated literature allows for empirical study.

* Address for correspondence: Glyn Lewis, Ph.D., Professor of Psychiatric Epidemiology, University of Bristol, Cotham House, Cotham Hill, Bristol BS6 6JL, UK.
(Email: Glyn.Lewis@bristol.ac.uk)

The measurement of associations between deployment to the Gulf War and later symptomatology is hampered by the necessary reliance on self-report. All measurements of health outcomes are also prone to random error. However, such outcomes, which are difficult to quantify objectively, might be particularly prone both to random measurement error and more importantly to measurement or reporting bias. There have been a number of suggestions that physical exposures such as multiple vaccinations, depleted uranium and oil fires might have caused illness in Gulf War veterans. This study has not focused on this question because most research into illness in Gulf War veterans began some years after the conflict ended and so it was difficult to establish exposure data without introducing bias.

This paper aims to identify and summarize the findings from studies that have compared the prevalence of chronic fatigue syndrome, multiple chemical sensitivity, Centers for Disease Control and Prevention (CDC)-defined chronic multi-symptom illness, fibromyalgia or the specific individual symptoms of fatigue and numbness in Gulf War veterans and an unexposed comparison group. We further aimed to critically assess the methodological quality of the individual studies. We have chosen to carry out a systematic review because of the advantages of this methodology in reducing bias and aiding interpretation of the findings (Lewis *et al.* 1997).

METHOD

Searching

The methods employed in the systematic review have been described in detail in another paper (Stimpson *et al.* 2003) and are merely summarized here. In total, 5387 studies from the period January 1990 to May 2001 were identified for possible inclusion by searching through databases (including EMBASE, Medline, ASSIA, SIGLE, PsycINFO, CancerLit and HealthStar, Current Contents, CINAHL and Biological Abstracts) and websites and by contacting researchers in the field. Studies were eligible for inclusion if they contained data on military, medical or peace-keeping personnel who were deployed to the Gulf War together with a comparison group which differed in its level of exposure. Abstracts of 2296 references that remained eligible were examined by two members of the research team. Studies were excluded if they measured simulated exposures, if they measured non-health-related outcomes or if the subjects were inhabitants of the Persian Gulf rather than deployed military personnel.

All identified papers that fulfilled the prestated inclusion criteria were categorized by health outcome. Fifty-six studies entered the category of symptomatology. Since the task of systematically reviewing the association between any symptom and any exposure would be too great and would not provide a digestible summary, we decided to focus on a limited number of important outcomes. These specific outcomes were chronic fatigue syndrome, multiple chemical sensitivity, CDC-defined chronic multi-symptom illness, fibromyalgia, or symptoms of either fatigue or numbness and tingling. Eighteen of the 56 papers did not include any of these outcomes. We further excluded those references that only compared exposure histories between Gulf War veterans ($n=8$), and studies which either presented data as a factor analysis or which included an inappropriate control group for the purposes of this review, e.g. comparison of ill Gulf War veterans and healthy community controls ($n=14$).

An updated search of the literature from January 2001 to May 2004 was completed which identified 538 references. Of databases searched in 2001, CancerLit and HealthStar are now incorporated into Medline and Dissertation Abstracts & Health and Psychosocial Instruments are not available. Two databases *not* searched in 2001, the Web of Knowledge Databases, Science Citation Index and Social Science Citation Index were included in this updated search. Of 538 potentially relevant references, 105 were selected from the abstract (or title if no abstract) that could be research studies with a comparison group. Seven papers that fulfilled the inclusion criteria and which contained data relating to the specified outcomes were identified. Therefore this review summarizes the findings from 23 references which compared Gulf War veterans with non-Gulf veterans (Sutker *et al.* 1993; Iowa Persian Gulf Study Group, 1997; Bell *et al.* 1998; Fukuda *et al.* 1998; Goss Gilroy Inc., 1998; Proctor *et al.* 1998, 2001,

Table 1. *Characteristics of studies that have investigated the association between deployment to the Gulf War and multi-symptom conditions amongst veterans*

First-named author	Study design	Sample	Study period	Ascertainment	Confounding
Sutker, 1993	Questionnaire administered by VA staff	215 GWV 44 NGV Sampled from 5 National Guard and Army Reserve units as part of debriefing programme	4–10 months after return from Operation Desert Shield/Storm	70% GWV (306 eligible) Number of eligible NGVs unclear	
Iowa Group, 1997, (Black, 2000), (Doebbeling, 2000)	Cross-sectional telephone interview survey	1896 GWV 1799 NGV DoD Manpower Data Center used to create a stratified random sample from 28968 military personnel from Iowa	Sept. 1995–May 1996	78·3% GWV (2421 eligible 73% NGV (2465 eligible) Those who were regulars, enlisted, navy or coast guard, aged up to 25 years and black or other ethnic background less likely to participate	Analysis accounted for sampling design
Bell, 1998	Cross-sectional telephone survey	24 GWV 17 NGV Random sample of 100 veterans enrolled at Tucson VA medical center	Autumn 1996–Spring 1997	759 GWV eligible 1756 NGV eligible 86% GWV (24/28) 3% of total eligible 85% NGV (17/20) 1% of total eligible	
Fukuda, 1998	Cross-sectional survey	1163 GWV 2560 NGV Sampled from 2 Air National Guard units Pennsylvania, 1 Air Force Reserve unit Florida, 1 active duty air force unit Florida	Jan.–Mar. 1995	6151 eligible Unit response rate 35–73%	
Goss Gilroy, 1998	Cross-sectional survey	3113 Canadian GWV 3439 Canadian NGV Sampled from Department of National Defence human resources data files	June–Sept. 1997	73·0% GWV (4262 eligible) 60·3% NGV (5699 eligible)	NGVs matched on sex, age group and regular/reserve status
Proctor, 1998	Cohort study	220 GWV from Fort Devens 73 GWV from New Orleans 50 Germany deployed veterans	Spring 1994–Autumn 1996	51% Fort Devens (353 sampled, 2949 eligible) 38% New Orleans (194 sampled, 928 eligible) 51% Germany deployed	Analysis accounted for sampling design
Proctor, 2001	As Proctor, 1998	180 GWV (Fort Devens) 46 Germany deployed veterans	As Proctor, 1998	As Proctor, 1998	As Proctor, 1998

[*continues overleaf*

Table 1 (*cont.*)

First-named author	Study design	Sample	Study period	Ascertainment	Confounding
Gray, 1999	Cross-sectional survey	527 GWV 970 NGV All active-duty Seabees who remained in service after ODS and were serving in 1 of 2 large Seabee Centres (California, Mississippi)	Sept. 1994–June 1995	Total eligible unknown Unit response rate 26–71%	
Ishoy, 1999, (Proctor, 2003)	Cross-sectional survey	686 GWV 231 NGV Selected from Danish Armed Forces personnel database. Proctor (2003) used subset of original study: 143 GWV 72 NGV	Feb. 1997–Jan. 1998 Feb.–Dec. 1999	83·6% GWV (821 eligible) 57·8% NGV (400 sampled) 75% response	Controls matched for gender, age and profession
Unwin, 1999, (Reid, 2001)	Cross-sectional postal survey	2735 GWV 2393 Bosnia veterans 2422 Era controls Random sample taken from the UK MoD database Unwin: servicemen only Reid: servicemen and women	Aug. 1997–Nov. 1998	70·4% GWV (4250 eligible) 61·9% Bosnia (4250 eligible) 62·9% Era (4246 eligible)	Sample was stratified by service, status, sex, age, rank and fitness
Unwin, 2002	As Unwin, 1999	236 GWV 217 Bosnia veterans 192 Era controls Servicewomen only	As Unwin, 1999	74·0% GWV 70·2% Bosnia 69·6% Era	As Unwin, 1999
Hotopf, 2003	As Unwin, 1999	1089 GWV 638 Bosnia veterans 643 Era controls	4-year follow-up (2001)	71·7% GWV 66·2% Bosnia 58·4% Era Non-responders to follow-up were more likely to be male, younger, unmarried and report poorer physical health at baseline	As Unwin, 1999. Follow-up sample stratified by baseline fatigue
Kang, 2000, (Kang, 2003)	Population-based survey	11441 GWV 9476 NGV Stratified random sample taken from the DoD Manpower Data Center	1995–1997	76·3% GWV (15000 eligible) 63 2% NGV (15000 eligible) Non-responders were more likely to be younger, unmarried, non-white and enlisted	Analysis accounted for sampling design
Steele, 2000	Population-based telephone survey	1548 GWV 482 NGV DoD Manpower Data Center used to create a stratified random sample from 16566 military personnel from Kansas	Feb.–Aug. 1998	65% response (3138 eligible) Overall, Gulf and female veterans were more likely to respond	

2003; Gray *et al.* 1999; Ishoy *et al.* 1999; Unwin *et al.* 1999; Black *et al.* 2000; Doebbeling *et al.* 2000; Kang *et al.* 2000, 2003; Steele, 2000; Cherry *et al.* 2001; Reid *et al.* 2001; White *et al.* 2001; Gray *et al.* 2002; Unwin *et al.* 2002; Hotopf *et al.* 2003; Simmons *et al.* 2004). These non-Gulf veterans may or may not have been deployed elsewhere on active duty.

Data extraction

Data relating to the studies' main hypotheses and to methodological quality were extracted independently by two members of the research team onto pre-designed data extraction forms. Some studies did not provide sufficient data to provide odds ratios (ORs) that were used as our outcome measure. Information on methodological quality of the individual studies included the statistical power, the potential of selection bias in the sampling of subjects and in the response rate, the potential bias in the measurement of outcomes, and the availability of data on confounders and the controlling for such variables in each study. We did not summarize the quality by creating a score as this approach has been criticized as giving equal weight to different aspects of methodology and relying heavily upon the quality of reporting rather than the quality of the original study (Juni *et al.* 2001).

Statistical analysis

A summary OR was calculated with a random-effects model using the DerSimonian & Laird method (1986). The estimate of heterogeneity between studies was taken from the inverse variance fixed-effects model. All analyses were performed using the 'metan' (Bradburn *et al.* 1998) command in STATA version 6 (Stata Corporation, College Station, TX, USA).

We chose to use this approach because of our *a priori* view that the studies were inherently heterogeneous. Using a ratio measure should reduce the heterogeneity resulting from different thresholds of case-definitions. A random-effects model assumes that the studies in a meta-analysis are randomly sampled from a distribution of effect sizes so the procedure estimates the mean of this distribution. In contrast, a fixed-effects model assumes that all the studies are sampled from a population with a single

Cherry, 2001	Cross-sectional survey	8210 GWV 3981 NGV Random sample taken from the UK MoD database	Dec. 1997–Sept. 1999	86·4% GWV (9505 eligible) 83·9% NGV (4749 eligible) Non-responders were younger	Sample was stratified by sex, age, service and rank
Gray, 2002	Cross-sectional survey	3831 GWV 3104 non-deployed All active-duty Seabees regardless of whether still in service at time of study identified through Defense Manpower Data Center 4933 deployed elsewhere	May 1997–May 1999	Total eligible 18945 Respondents more likely to be reservists, married, Caucasian, and deployed elsewhere Overall response rate 63 6%	GWV more likely to be reservists, male, younger, unmarried, have more evidence of cognitive failure and report more days lost to illness in last 12 months
White, 2001	Cross-sectional survey	193 GWV 47 Germany deployed Subset of Proctor (11)	Spring 1995	62% response rate from within a survey that had a 78% response rate. Overall response 48%. Active duty and African–American service personnel over-represented	Adjusted for age, education and gender
Simmons, 2004	Cross-sectional survey	23358 GWV 17730 NGV	Aug. 1998–March 2001	48% response rate. Analysis restricted to men	Adjusted for age, service, rank, serving status at time of survey, smoking and current alcohol consumption.

References within parentheses are based on identical samples. GWV, Gulf War veterans; NGV, non-Gulf veterans.

effect size which all studies are estimating. There was insufficient data to further investigate heterogeneity.

RESULTS

We identified 23 references to 14 primary study samples which compared the prevalence of multi-symptom conditions in both Gulf War veterans and non-Gulf veterans. Black *et al.* (2000) and Doebbeling *et al.* (2000) utilized the sample described in the Iowa study (Iowa Persian Gulf Study Group, 1997); the study sample described by Proctor *et al.* (2001) and White *et al.* (2001) was a smaller subset of that described by Proctor *et al.* (1998); the sample initially described by Unwin *et al.* (1999; servicemen only) was later utilized by Reid *et al.* (2001; servicemen and women), Unwin *et al.* (2002; servicewomen only) and Hotopf *et al.* (2003; 4-year follow-up study); Kang *et al.* reported data on the same study sample both in 2000 and in 2003. Although both publications by Gray *et al.* (1999, 2002) included a sample of active-duty Seabees, these were separate studies with different sampling strategies and fieldwork dates.

Table 1 summarizes the methodology used in creating these 14 study samples. Nine samples consisted of American Gulf War veterans, three samples included British veterans, one sample included Canadian veterans and one sample included Danish veterans involved in UN peacekeeping and humanitarian operations. Sutker *et al.* (1993) were the only authors to assess multi-symptom conditions soon after returning from the war arena. All other studies assessed these conditions from 1994 onwards. All studies relied on self-reported measurement of symptoms. Three studies investigated the test–retest reliability of their measures. The Iowa Group (1997) reported kappas between 0·35 and 0·79. Gray *et al.* (1999) retested a subsample of subjects between 5 and 9 months later and reported kappas varying between 0·01 and 0·86. Gray *et al.* (2002) also tested the reliability of self-reported physician-diagnosed medical conditions in 519 subjects retested 6 months apart and reported a kappa of 0·6. These studies cannot investigate the issue of bias and their measure of reliability will be underestimates as symptoms may also have changed in the intervening period.

Table 2 summarizes the results of the 10 references that investigated the association between Gulf deployment and chronic fatigue. There was variation in the length of period prevalence estimated and in the diagnosis of chronic fatigue. Data from seven of the references were entered into the meta-analysis. The summary OR reflected a statistically significant association between deployment to the Gulf War and subsequent reporting of chronic fatigue [OR 3·81, 95% confidence interval (CI) 2·17–6·71]. The test for heterogeneity was statistically significant ($p < 0{\cdot}001$).

We identified 16 references (Table 3) that investigated the association between Gulf deployment and symptoms of fatigue. Cherry *et al.* (2001), White *et al.* (2001) and Proctor *et al.* (2003) did not quote prevalence estimates of symptoms, therefore their data cannot provide ORs. Only the study by Sutker *et al.* (1993) carried out within 10 months of returning from the war reported a non-significant increase in reporting of fatigue, although the CI was wide (OR 1·98, 95% CI 0·73–5·38). All other studies (Fukuda *et al.* 1998; Gray *et al.* 1999; Ishoy *et al.* 1999; Unwin *et al.* 1999, 2002; Doebbeling *et al.* 2000; Kang *et al.* 2000; Steele, 2000; Proctor *et al.* 2001; Gray *et al.* 2002; Hotopf *et al.* 2003; Simmons *et al.* 2004) reported a statistically significant association between deployment and symptoms of fatigue. This was summarized by an OR of 3·74 (95% CI 2·94–4·75) when 12 studies were entered into a meta-analysis using a random-effects model. The test for heterogeneity based on the fixed-effects model was statistically significant ($\chi^2 = 59{\cdot}4$, df = 6, $p < 0{\cdot}001$), probably due to the difference in effect sizes demonstrated between Unwin *et al.* (1999), Kang *et al.* (2000) and Gray *et al.* (2002). It was striking that the prevalence of self-reported fatigue was very high amongst British Gulf war veterans, Bosnia veterans and era controls in the study by Unwin *et al.* (1999, 2002). Hotopf *et al.* (2003) demonstrated that 4 years later the Gulf War veterans continued to report more fatigue than the comparison military cohorts. This seemed to be largely explained by greater persistence of fatigue symptoms amongst Gulf veterans, but there was also some evidence of higher incidence of fatigue

Table 2. *Studies that have investigated the association between deployment to the Gulf War and chronic fatigue syndrome amongst veterans*

First-named author	Outcome definition	Prevalence	OR (95% CI) in relation to Gulf deployment	% weight in meta-analysis
Iowa Group, 1997	Centers for Disease Control and Prevention (CDC) Chronic Fatigue Syndrome Questionnaire and Chalder *et al.*'s (1993) Fatigue Scale. Prevalence 1 month prior to interview	Regular military GWV prevalence 1·0% NGV prevalence 0·2% National Guard/Reserves GWV prevalence 2·9% NGV prevalence 1·1%	3·15 (1·60–6·20)	15·0
Goss Gilroy, 1998	CDC Chronic Fatigue Syndrome Questionnaire. Period of prevalence is unclear	OR chronic fatigue 5·27 (95% CI 3·95–7·03)		
Unwin, 1999	Chalder *et al.*'s Fatigue Scale. Prevalence 1 month prior to interview	GWV prevalence 3·3% Bosnia prevalence 0·8% Era prevalence 0·8%		
Reid, 2001	CDC diagnosis derived from Chalder *et al.*'s Fatigue Scale and SF-36. Prevalence 1 month prior to interview	GWV prevalence 2·1% Era prevalence 1·8% Bosnia prevalence 0·7%	1·17 (0·81–1·69)	17·7
Unwin, 2002	Chalder *et al.*'s Fatigue Scale. Prevalence 1 month prior to interview	GWV prevalence 4·7% Bosnia prevalence 0·5% Era prevalence 0·5%		
Steele, 2000	CDC diagnosis derived from self-reported symptoms and fatigue characteristics. Prevalence 1 year prior to interview	GWV prevalence 7·1% NGV prevalence 0·7%	11·04 (3·49–34·93)	10·7
Proctor, 2001	CDC diagnosis derived from self-reported symptoms. Unknown prevalence period	GWV prevalence 7·5% (excluding rule-out indicators 2·0%) Germany deployed prevalence 0%	2·37 (0·13–44·83)	3·1
Gray, 2002	Self-reported physician diagnosis of chronic fatigue syndrome, onset after Aug. 1991	GWV prevalence 5·17% Deployed elsewhere prevalence 0·79% Non-deployed prevalence 0·68%	6·75 (4·77–9·54)	17·9
Kang, 2003	CDC diagnosis derived from self-reported symptoms. Prevalence 1 year and 1 month prior to interview	GWV prevalence 5·6% NGV prevalence 1·2%	4·87 (5·96–21·92)	18·7
Simmons, 2004	Prevalence of self-reported chronic fatigue syndrome since 1990	GWV 0·4% NGV 0·1%	3·5 (2·1–5·5)	16·9
			Summary OR 3·81 (2·17–6·71) Heterogeneity $\chi^2 = 59{\cdot}4$, df = 6, $p < 0{\cdot}0001$	

GWV, Gulf War veterans; NGV, non-Gulf veterans.

OR, Odds ratio with confidence intervals in parentheses. Summary OR derived from random-effects model.

Iowa study combines prevalence in regular military and National Guard/Reserves for meta-analysis.

Goss Gilroy not used in meta-analysis since raw numbers not provided. Unwin data not used in meta-analysis to avoid duplication with Reid data. Reid study compares Gulf *versus* era controls for meta-analysis. Proctor study uses prevalence excluding rule-out indicators for meta-analysis. Gray study compares Gulf *versus* controls deployed elsewhere for meta-analysis.

in the two deployed groups when compared to the era controls.

Seven references investigated the association between deployment to the Gulf and multiple chemical sensitivity, five of which provided data for the meta-analysis with the study by Black *et al.* (2000) and Gray *et al.* (2002) providing approximately 65% of the weight to the summary OR (Table 4). The prevalence estimates quoted by Bell *et al.* (1998) were much higher than those reported in the other studies probably due to the much broader

Table 3. *Studies that have investigated the association between deployment to the Gulf War and symptoms of fatigue amongst veterans*

First-named author	Outcome definition	Prevalence	OR in relation to Gulf deployment	% weight in meta-analysis
Sutker, 1993	Current fatigue/lack of energy	GWV prevalence 20% NGV prevalence 11%	1·98 (0·73–5·38)	3·7
Fukuda, 1998	Current fatigue	GWV prevalence 43% NGV prevalence 17%	3·69 (3·16–4·31)	9·9
Gray, 1999	Unusual fatigue ⩾1 month since July 1990	GWV prevalence 20·1% NGV prevalence 5·4%	4·44 (3·13–6·32)	8·5
Ishoy, 1999	Abnormal feeling of fatigue during preceding 12 months	Symptom after Aug. 1990 GWV prevalence 26·4% NGV prevalence 10·8% Symptom before Aug. 1990 GWV prevalence 1·9% NGV prevalence 1·7%	2·95 (1·89–4·62)	7·6
Unwin, 1999	Current fatigue (past month)	GWV prevalence 50·7% Bosnia prevalence 26·3% Era prevalence 27·7%	2·68 (2·40–3·00)	10·1
Unwin, 2002	Current fatigue (past month)	GWV prevalence 59·4% Bosnia prevalence 31·6% Era prevalence 34·7%	1·9 (1·1–3·2)	8·1
Hotopf, 2003	Current fatigue	GWV prevalence 43·4% Bosnia prevalence 32·7% Era prevalence 22·0%		
Doebbeling, 2000	Period of extreme fatigue every day or almost every day in past year	GWV prevalence 23% NGV prevalence 9%	3·02 (2·49–3·66)	9·7
Kang, 2000	Fatigue Period of prevalence unclear	GWV prevalence 38% NGV prevalence 15%	3·47 (3·25–3·72)	10·3
Steele, 2000	Period of fatigue in past year	GWV prevalence 36% NGV prevalence 12%	4·14 (3·01–5·68)	8·7
Proctor, 2001	Current fatigue not meeting CDC diagnosis	GWV prevalence 29·4% Germany deployed prevalence 8·7%	4·38 (1·50–12·83)	3·4
Cherry, 2001	Tiredness (past month) Visual analogue scale	GWV mean score = 8·3 NGV mean score = 5·0		
Gray, 2002	Unusual fatigue in past year	GWV prevalence 38·95% Deployed elsewhere prevalence 14·13% Non-deployed prevalence 13·43%	3·87 (3·50–4·29)	10·1
White, 2001	Profile of Mood States	GWV *t* score 47·4, S.E. = 0·62 NGV *t* score 42·1, S.E. = 0·84 Difference $p = 0·001$		
Simmons, 2004	Self-reported general fatigue	GWV prevalence 10·8% NGV prevalence 1·2%	9·6 (8·3–11·1)	10·0
Proctor, 2003	Profile of Mood States (Raw scores)	GWV 6·3, S.E. = 0·4 NGV 4·6, S.E. = 0·5 Difference $p = 0·005$		
			Summary OR 3·74 (2·94–4·75) Heterogeneity $\chi^2 = 241·8$, df = 11, $p < 0·0001$	

GWV, Gulf War veterans; NGV, non-Gulf veterans.

OR, Odds ratio with confidence intervals in parentheses. Summary OR derived from random-effects model.

Unwin studies compare Gulf *versus* era controls for meta-analysis. Hotopf data not used in meta-analysis to avoid duplication with Unwin data. Gray study compares Gulf *versus* controls deployed elsewhere for meta-analysis.

definition of the outcome. Overall, the odds of reporting multiple chemical sensitivity was approximately three and a half times greater amongst veterans deployed to the Gulf War (OR 3·56, 95% CI 2·03–6·24). The test for heterogeneity was of borderline statistical significance ($p = 0·05$).

All five studies that investigated chronic multi-symptom illness used the CDC diagnosis (Table 5). Prevalence estimates varied greatly

Table 4. *Studies that have investigated the association between deployment to the Gulf War and multiple chemical sensitivity amongst veterans*

First-named author	Outcome definition	Prevalence	OR (95% CI) in relation to Gulf deployment	% weight in meta-analysis
Bell, 1998	Consider self especially sensitive to certain chemicals currently	GWV prevalence 62·5% NGV prevalence 41·2%	2·38 (0·67–8·49)	13·1
Goss Gilroy, 1998	Report illness from chemical sensitivity, ⩾2 incitants, ⩾2 organ system symptoms, behaviour change	OR multiple chemical sensitivity by Gulf deployment 4·01 (95% CI 2·43–6·62)		
Unwin, 1999	Shortened version of Kipen *et al.*'s (1995) measurement of symptoms of possible chemical sensitivity	GWV prevalence 0·8% Bosnia prevalence 0·4% Era prevalence 0·3%		
Reid, 2001	Symptoms reported ⩾3 organ sites including CNS, duration ⩾3 months, sensitivity to ⩾4 substances from shortened version of Kipen *et al.*'s (1995) list	GWV prevalence 1·3% Bosnia prevalence 0·3% Era prevalence 0·2%	6·89 (2·73–17·36)	19·3
Black, 2000	Report illness from chemical sensitivity, ⩾2 incitants, ⩾2 organ system symptoms, impairment or behaviour change	GWV prevalence 5·41% NGV prevalence 2·55%	2·19 (1·54–3·12)	35·0
Proctor, 2001	>1 symptom that began after service, ⩾2 organ system symptoms, symptoms triggered by low-level chemical exposure, symptoms that cannot be explained by any diagnosis	GWV prevalence 2·9% Germany deployed prevalence 0%	2·91 (0·16–53·66)	3·4
Gray, 2002	Self-reported physician diagnosis of multiple chemical sensitivity, onset after Aug. 1991	GWV prevalence 1·62% Deployed elsewhere prevalence 0·32% Non-deployed prevalence 0·39%	5·06 (2·91–8·77)	29·2
			Summary OR 3·56 (2·03–6·24) Heterogeneity $\chi^2=9·7$, df = 4, $p=0·05$	

GWV, Gulf War veterans; NGV, non-Gulf veterans.

OR, Odds ratio with confidence intervals in parentheses. Summary OR derived from random-effects model.

Goss Gilroy data not used in meta-analysis since raw numbers not provided. Unwin data not used in meta-analysis to avoid duplication with Reid data. Reid study compares Gulf *versus* era controls for meta-analysis. Gray study compares Gulf *versus* controls deployed elsewhere for meta-analysis.

depending on the severity of the illness. For those studies that reported prevalence of severe and moderate illness separately the estimates were summed to provide one OR for each study. Although Gulf deployment was strongly associated with chronic multi-symptom illness (OR 3·62, 95% CI 2·75–4·76), the test for heterogeneity between studies was highly statistically significant ($p<0·001$).

There were six studies which reported prevalence of numbness or tingling sensations (Proctor *et al.* 1998; Fukuda *et al.* 1998; Ishoy *et al.* 1999; Unwin *et al.* 1999; Kang *et al.* 2000; Cherry *et al.* 2001). Deployment to the Gulf War was associated with a two-fold increase in odds of reporting such sensations (OR 2·37, 95% CI 1·80–3·12). The variation in definition of outcome and period of prevalence across studies might have led to the statistically significant heterogeneity ($\chi^2=35·5$, df = 4, $p<0·001$).

Only two studies reported the association between Gulf deployment and fibromyalgia. The Iowa study (1997) reported a 1-month prevalence of 18·2% amongst Gulf War regular military veterans and a prevalence of 9·2% amongst non-Gulf regular military veterans. Goss Gilroy (1998) did not quote prevalence figures, but reported an OR of 1·81 (95% CI 1·55–2·13) with respect to Gulf deployment.

DISCUSSION

This systematic review supports the hypothesis that deployment to the Gulf War is associated with greater reporting of multi-symptom

Table 5. *Studies that have investigated the association between deployment to the Gulf War and defined 'chronic multi-symptom illness' amongst veterans*

First-named author	Outcome definition	Prevalence	OR (95% CI) in relation to Gulf deployment	% weight in meta-analysis
Fukuda, 1998	CDC Chronic multi-symptom illness: ⩾1 symptom (6 months) from ⩾2 of following: fatigue, mood and cognition, musculoskeletal	Severe GWV prevalence 6·0% NGV prevalence 0·7% Mild to moderate GWV prevalence 39% NGV prevalence 14%	4·76 (4·05–5·58)	24·8
Unwin, 1999	CDC Chronic multi-symptom illness: ⩾1 symptom (6 months) from ⩾2 of following: fatigue, mood and cognition, musculoskeletal	GWV prevalence 61·9% Bosnia prevalence 36·8% Era prevalence 36·4%	2·84 (2·54–3·16)	26·0
Unwin, 2002	CDC Chronic multi-symptom illness	GWV prevalence 63·8% Bosnia prevalence 36·6% Era prevalence 34·7%	3·33 (2·23–4·96)	17·3
Steele, 2000	CDC Chronic multi-symptom illness: ⩾1 symptom (6 months) from ⩾2 of cognition, musculoskeletal	GWV prevalence 47·2% NGV prevalence 19·8%	3·63 (2·81–4·68)	21·9
Proctor, 2001	CDC Chronic multi-symptom illness: ⩾1 symptom (6 months) from ⩾2 of following: fatigue, mood and cognition, musculoskeletal	GWV prevalence 65·3% Germany deployed prevalence 32·6% Severe GWV prevalence 31·9% NGV prevalence 4·4% Mild to moderate GWV prevalence 33·4% NGV prevalence 28·3%	3·93 (1·98–7·83)	10·0
			Summary OR 3·62 (2·75–4·76) Heterogeneity $\chi^2 = 28{\cdot}0$, df = 4, $p < 0{\cdot}001$	

GWV, Gulf War veterans; NGV, non-Gulf veterans.
OR, Odds ratio with confidence intervals in parentheses. Summary OR derived from random-effects model.
Fukuda study sums prevalence of severe and mild-moderate symptoms for meta-analysis. Unwin study compares Gulf *versus* era controls for meta-analysis. Proctor study sums prevalence of severe and mild-moderate symptoms for meta-analysis.

conditions. Each of the primary studies consistently demonstrated that a greater proportion of Gulf war veterans than non-Gulf veterans reported either a defined multi-symptom condition or individual symptoms of fatigue and numbness. Gulf deployment was most strongly associated with chronic fatigue, with Gulf War veterans being almost four times more likely than non-Gulf veterans to report such symptoms. The odds of reporting multiple chemical sensitivity or chronic multi-symptom illness as defined by CDC were also approximately three and a half times greater amongst Gulf War veterans. The evidence for an association with fibromyalgia was limited to the results from only two studies.

We used a random-effects model in the meta-analyses to estimate summary ORs. This procedure estimates the mean of the distribution of possible effect sizes. We also quote a test for heterogeneity based on a fixed-effects model. There is strong evidence that there is not a single effect size underlying the variation in results of these studies. There are a number of reasons why this should be so, including variation in the case-definition used in the different studies as well as the difference in sampling and between the nations. However, all the studies reported an association between Gulf deployment and multi-symptom conditions so the heterogeneity arises from difference in strength rather than direction of association.

Irrespective of the relative risk, the prevalence of chronic fatigue and multiple chemical sensitivity amongst Gulf War veterans was relatively low (maximum estimate approximately 7%). The prevalence of CDC-defined chronic multi-symptom illness was much higher amongst Gulf veterans, but was also reported in over 30% of

veterans who had not been deployed to the Gulf arena.

Limitations of primary studies

Sample size

The number of Gulf War veterans sampled ranged from 24 (Bell *et al.* 1998) to over 23 000 (Simmons *et al.* 2004). This variation in sample size can be seen in the differential weighting of studies in the meta-analysis. Although a small sample size might be a limitation of an individual study, the possibility of a Type II error occurring in this review is minimized by the pooling of data from all appropriate studies to increase statistical power.

Sample selection and response bias

In a cross-sectional survey it is important to derive a random sample of all those subjects who are potentially eligible in order to generate a representative sample of the larger population of interest. Those studies which selected a random sample of veterans from either American, British, Canadian or Danish military personnel databases are likely to have fulfilled this criterion. However, those studies which sampled from individual military units are more prone to selection bias (Sutker *et al.* 1993; Bell *et al.* 1998; Fukuda *et al.* 1998; Proctor *et al.* 1998, 2001; Gray *et al.* 1999).

Furthermore, once the sample has been selected it is important to maintain its ability to represent the larger population by achieving a high response rate. In general, most of the studies achieved a satisfactory response rate of $\geqslant$70% amongst veterans of the Gulf War. However the response rate amongst non-Gulf veterans tended to be systematically lower in most studies for which data were available. Unwin *et al.* (1999) intensively followed up a random selection of non-responders and found that those with more symptoms responded earlier, although mean SF-36 ratings of health perception and ratings of physical function did not differ significantly by response time. There was no significant interaction between deployment, late response and health outcome since Gulf War late responders did not differ from Bosnia or era late responders. So the prevalence estimate of symptoms might be a biased overestimate, but relative measures of effect, as reported here, should be less prone to bias.

Measurement of health outcomes

All of the studies relied on the veterans' self-report of symptoms without any alternative method of ascertainment. This is prone both to random measurement error which would tend to nullify any association between Gulf deployment and symptom reporting, and more importantly to measurement bias. All the symptoms and, therefore, case-definitions rely upon self-report and independent ascertainment is impossible. Gray *et al.* (2002) quoted a kappa coefficient of 0·6 indicating moderate reliability of self-reported physician-diagnosed medical conditions but reliability was poorer (kappa $-0{\cdot}01$ to 0·86) for self-reported symptoms in 139 subjects. It is noteworthy that only Sutker *et al.* (1993) measured the prevalence of fatigue symptoms soon after the veterans' return from the Gulf whereas the remainder of studies measured symptoms after 1994, by which time media interest in the health of Gulf War veterans might have led to an increased awareness and increased reporting of symptoms.

It is worth mentioning that the diagnostic criteria for the multi-symptom conditions that are included in this review remain controversial, and indeed it has been suggested that these 'functional somatic illnesses' might be variants of one another (Wessely *et al.* 1999). Some might argue that the similarity of ORs for the multi-symptom conditions supports the idea that they have a common aetiology. However, the controversial nature of these syndromes should not affect our conclusions as the studies included in the review have used the same case-definition when comparing Gulf War Veterans with those not deployed to the Gulf.

Confounding

The adjustment for confounders varied between studies (see Table 1) but the later and larger studies, which also contributed more to the meta-analysis, tended to deal with this more throughly. The studies of Goss Gilroy (1998), Ishoy *et al.* (1999), Unwin *et al.* (1999) and Cherry *et al.* (2001) used veterans matched on age, sex and some aspect of military status. An important confounder for these health

outcomes would be symptom reporting before deployment to the Gulf. Ishoy *et al.* (1999) and White *et al.* (2001) asked veterans about symptoms with an onset after August 1990, but the questionnaires were administered in the late 1990s so the responses are still open to recall bias. Ishoy *et al.* (1999) reported that the prevalence of symptoms of fatigue with onset before August 1990 was very similar amongst Gulf and non-Gulf veterans and was much lower than the prevalence of symptoms with onset after the Gulf. This would suggest that the association between Gulf deployment and symptoms of fatigue is not confounded by baseline symptom severity.

In general, adjustment for confounders in these studies tended to reduce the strength of association between Gulf deployment and multi-symptom conditions. The results for the meta-analysis are on unadjusted findings so are likely to overestimate the association. On the other hand, the idea that deployed troops were healthier than their non-deployed counterparts is also quite plausible, at least for the US army, in which case the inability to adjust for this might suggest that the reported findings are underestimates.

Strengths and limitations of this review

This review benefits from a highly sensitive search strategy based on both published material and on grey literature such as conference abstracts and preliminary reports. Inclusion and exclusion criteria were independently assessed by two reviewers and in the case of any disagreement the reference was included until the next stage of data extraction to ensure no references were excluded unnecessarily. However, as in all systematic reviews, we may have failed to identify some studies.

This review has been limited to investigating associations between Gulf deployment *versus* non-deployment and reporting of various symptoms. We cannot offer any insight into the possible mechanisms underlying the observed associations. It is also possible that an association might exist between a specific environmental exposure (e.g. threat of chemical warfare agents, pyridostigmine bromide, petrochemicals, non-routine immunizations) and reporting of symptoms. We chose not to examine these associations in this review due to the problems associated with the inaccuracy of such self-reported exposures.

CONCLUSION

We identified 23 references that compared the prevalence of multi-symptom conditions in both Gulf War veterans and non-Gulf veterans. The results support the hypothesis that deployment to the Gulf War is associated with greater reporting of multi-symptom conditions. Gulf deployment was most strongly associated with chronic fatigue, with Gulf War veterans being almost four times more likely than non-Gulf veterans to report such symptoms. However it seems likely that at least some of this association might be explained by response bias, measurement bias and confounding. This review cannot shed light on any biological or socio-cultural mechanisms underlying this association but adds to the increasing body of evidence that suggests veterans of military deployment are more likely to subsequently report a variety of poorly understood complaints including fatigue, pain and associated symptoms.

ACKNOWLEDGEMENTS

We thank the Information Services Department at the University of Wales College of Medicine for assistance in obtaining references for the review. This project was funded by the Ministry of Defence through the Medical Research Council.

DECLARATION OF INTEREST

None.

REFERENCES

Bell, I. R., Warg-Damiani, L., Baldwin, C. M., Walsh, M. E. & Schwartz, G. E. (1998). Self-reported chemical sensitivity and wartime chemical exposures in Gulf War veterans with and without decreased global health ratings. *Military Medicine* **163**, 725–732.

Black, D. W., Doebbeling, B. N., Voelker, M. D., Clarke, W. R., Woolson, R. F., Barrett, D. H. & Schwartz, D. A. (2000). Multiple chemical sensitivity syndrome – Symptom prevalence and risk factors in a military population. *Archives of Internal Medicine* **160**, 1169–1176.

Bradburn, M. J., Deeks, J. J. & Altman, D. G. (1998). Sbe24: metan – an alternative meta-analysis command. *Stata Technical Bulletin* **44**, 4–15.

Chalder, T., Berelowitz, G., Pawlikowska, T., Watts, L., Wessely, S., Wright, D. & Wallace, E. P. (1993). Development of a fatigue scale. *Journal of Psychosomatic Research* **37**, 147–153.

Cherry, N., Creed, F., Silman, A., Dunn, G., Baxter, D., Smedley, J., Taylor, S. & Macfarlane, G. J. (2001). Health and exposures of United Kingdom Gulf war veterans. Part I: the pattern and extent of ill health. *Occupational and Environmental Medicine* **58**, 291–298.

DerSimonian, R. & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **17**, 177–188.

Doebbeling, B. N., Clarke, W. R., Watson, D., Torner, J. C., Woolson, R. F., Voelker, M. D., Barrett, D. H. & Schwartz, D. A. (2000). Is there a Persian Gulf War syndrome? Evidence from a large population-based survey of veterans and nondeployed controls. *American Journal of Medicine* **108**, 695–704.

Fukuda, K., Nisenbaum, R., Stewart, G., Thompson, W. W., Robin, L., Washko, R. M., Noah, D. L., Barrett, D. H., Randall, B., Herwaldt, B. L., Mawle, A. C. & Reeves, W. C. (1998). Chronic multisymptom illness affecting Air Force veterans of the Gulf War. *Journal of the American Medical Association* **280**, 981–988.

Goss Gilroy Inc. (1998). *Health Study of Canadian Forces Personnel Involved in the 1991 Conflict in the Persian Gulf*, vol. 1. Department of National Defence: Ontario, Canada.

Gray, G. C., Kaiser, K. S., Hawksworth, A. W., Hall, F. W. & Barrett-Connor, E. (1999). Increased postwar symptoms and psychological morbidity among U.S. Navy Gulf War veterans. *American Journal of Tropical Medicine and Hygiene* **60**, 758–766.

Gray, G. C., Reed, R. J., Kaiser, K. S., Smith, T. C. & Gastanaga, V. M. (2002). Self-reported symptoms and medical conditions among 11,868 Gulf War-era veterans. The Seabee Health Study. *American Journal of Epidemiology* **155**, 1033–1044.

Haley, R. W., Kurt, T. L. & Hom, J. (1997). Is there a Gulf War Syndrome? Searching for syndromes by factor analysis of symptoms. *Journal of the American Medical Association* **277**, 215–222.

Hotopf, M., David, A., Hull, L., Nikalaou, V., Unwin, C. & Wessely, S. (2003). Gulf war illness- better, worse, or just the same? A cohort study. *British Medical Journal* **327**, 1357–1358.

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

Ishoy, T., Suadicani, P., Guldager, B., Appleyard, M., Hein, H. O. & Gyntelberg, F. (1999). State of health after deployment in the Persian Gulf – The Danish Gulf War Study. *Danish Medicine Bulletin* **46**, 416–419.

Ismail, K., Everitt, B., Blatchley, N., Hull, L., Unwin, C., David, A. & Wessely, S. (1999). Is there a Gulf War syndrome? *Lancet* **353**, 179–182.

Juni, P., Altman, D. G. & Egger, M. (2001). Assessing the quality of randomised controlled trials. In *Systematic Reviews in Health Care* (ed. M. Egger, G. Davey Smith and D. G. Altman), pp. 87–108. BMJ Books: London.

Kang, H. K., Mahan, C. M., Lee, K. Y., Magee, C. A. & Murphy, F. M. (2000). Illnesses among United States veterans of the Gulf War: A population-based survey of 30,000 veterans. *Journal of Occupational and Environmental Medicine* **42**, 491–501.

Kang, H. K., Natelson, B. H., Mahan, C. M., Lee, K. Y. & Murphy, F. M. (2003). Post-traumatic stress disorder and chronic fatigue syndrome-like illness among Gulf War veterans: a population-based survey of 30,000 veterans. *American Journal of Epidemiology* **157**, 141–148.

Kipen, H. M., Hallman, W., Kelly-McNeil, K. & Fiedler, N. (1995). Measuring chemical sensitivity prevalence: a questionnaire for population studies. *American Journal of Public Health* **85**, 574–577.

Kroenke, K. & Price, R. (1993). Symptoms in the community: Prevalence, classification and psychiatric comorbidity. *Archives of Internal Medicine* **153**, 2474–2480.

Lewis, G., Churchill, R. & Hotopf, M. (1997). Systematic reviews and meta-analysis. *Psychological Medicine* **27**, 3–7.

Proctor, S. P., Heaton, K. J., White, R. F. & Wolfe, J. (2001). Chemical sensitivity and chronic fatigue in Gulf War veterans: a brief report. *Occupational and Environmental Medicine* **43**, 259–264.

Proctor, S. P., Heeren, T., White, R. F., Wolfe, J., Borgos, M. S., Davis, J. D., Pepper, L., Clapp, R., Sutker, P. B., Vasterling, J. J. & Ozonoff, D. (1998). Health status of Persian Gulf War veterans: self-reported symptoms, environmental exposures and the effect of stress. *International Journal of Epidemiology* **27**, 1000–1010.

Proctor, S. P., White, R., Heeren, T., Debes, F., Gloerfelt-Tarp, B., Appleyard, M., Ishoy, T., Guldager, B., Suadicini, P., Gyntelberg, F. & Ozonoff, D. M. (2003). Neuropsychological functioning in Danish Gulf War Veterans. *Journal of Psychopathology and Behavioural Assessment* **25**, 85–93.

Reid, S., Hotopf, M., Hull, L., Ismail, K., Unwin, C. & Wessely, S. (2001). Multiple chemical sensitivity and chronic fatigue syndrome in British Gulf War veterans. *American Journal of Epidemiology* **153**, 604–609.

Simmons, R., Maconochie, N. & Doyle, P. (2004). Self-reported ill health in male UK Gulf War veterans: a retrospective cohort. *BMC Public Health* **4**, 27.

Steele, L. (2000). Prevalence and patterns of Gulf War illness in Kansas veterans: association of symptoms with characteristics of person, place, and time of military service. *American Journal of Epidemiology* **152**, 992–1002.

Stimpson, N. J., Thomas, H. V., Weightman, A. L., Dunstan, F. & Lewis, G. (2003). Psychiatric disorder in veterans of the Persian Gulf War of 1991: systematic review. *British Journal of Psychiatry* **182**, 391–403.

Sutker, P. B., Uddo, M., Brailey, K. & Allain, A. N. (1993). War-zone trauma and stress-related symptoms in Operation Desert Shield/Storm (ODS) returnees. *Journal of Social Issues* **49**, 33–50.

Unwin, C., Blatchley, N., Coker, W., Ferry, S., Hotopf, M., Hull, L., Ismail, K., Palmer, I., David, A. & Wessely, S. (1999). Health of UK servicemen who served in Persian Gulf War. *Lancet* **353**, 169–178.

Unwin, C., Hotopf, M., Hull, L., Ismail, K., David, A. & Wessely, S. (2002). Women in the Persian Gulf: lack of gender differences in long-term health effects of service in United Kingdom Armed Forces in the 1991 Persian Gulf War. *Military Medicine* **167**, 406–413.

Wessely, S., Nimnuan, C. & Sharpe, M. (1999). Functional somatic syndromes: one or many? *Lancet* **354**, 936–939.

White, R. F., Proctor, S. P., Heeren, T., Wolfe, J., Krengel, M., Vasterling, J., Lindem, K., Heaton, K. J., Sutker, P. & Ozonoff, D. M. (2001). Neurpsychological function in Gulf War veterans: relationships to self-reported toxicant exposures. *American Journal of Industrial Medicine* **40**, 42–54.