

Psychiatric disorder in veterans of the Persian Gulf War of 1991

Systematic review

NICOLA J. STIMPSON, HOLLIE V. THOMAS, ALISON L. WEIGHTMAN, FRANK DUNSTAN and GLYN LEWIS

Background Veterans of the Persian Gulf War of 1991 have reported symptoms attributed to their military service.

Aims To review all studies comparing the prevalence of psychiatric disorders in Gulf War veterans and in a comparison group of service personnel not deployed to the Gulf War.

Method Studies of military personnel deployed to the Gulf published between 1990 and 2001 were identified from electronic databases. Reference lists and websites were searched and key researchers were contacted for information. A total of 2296 abstracts and 409 complete articles were reviewed and data were extracted independently by two members of the research team.

Results The prevalence of psychiatric disorder in 20 studies of Gulf War veterans was compared with the prevalence in the comparison group. Prevalence of post-traumatic stress disorder (PTSD) and common mental disorder were higher in the Gulf War veterans. Heterogeneity between studies was significant, but all reported this increased prevalence.

Conclusions Veterans of the Persian Gulf War reported an increased prevalence of PTSD and common mental disorder compared with other active service personnel not deployed to the Gulf. These findings are attributable to the increase in psychologically traumatic events in wartime.

Declaration of interest Funded by the UK Medical Research Council.

Since the end of the Persian Gulf War of 1991, its veterans have reported a range of health complaints attributed to service during the war. These veterans report an increased prevalence of a whole range of common symptoms compared with other service personnel who were not deployed to the Gulf.

It is now widely recognised that exposure to combat and other wartime experiences can have both short-term and long-term psychological effects. These psychological consequences are varied, but the concept of post-traumatic stress disorder (PTSD) has arisen to describe the syndrome of intrusive thoughts, flashbacks, hyperarousal and numbing that can occur after exposure to any traumatic event, including those common in wartime.

The Persian Gulf War was brief and there were relatively few casualties among the troops deployed on behalf of the United Nations. Nevertheless, a number of aspects of the war exposed service personnel to traumatic and stressful events: these included the risk of chemical and biological warfare, exposure to combat, and dealing with prisoners and dead and wounded Iraqi soldiers. This paper describes a systematic review of studies that have compared the prevalence of psychiatric disorder in Gulf War veterans with its prevalence in a comparison group who were not deployed to the Gulf (non-Gulf veterans).

METHOD

Data search

Studies between January 1990 and May 2001 were identified from a range of electronic databases, including EMBASE, Medline, ASSIA, SIGLE, PsycINFO, CancerLit, HealthSTAR, Dissertation, Abstracts, Current Contents, Health and Psychosocial Instruments, CINAHL and Biological Abstracts. Keywords used to identify the studies were: DESERT STORM or

DESERT SHIELD or DESERTSHIELD or GULF WAR or GULF SYNDROME or GULF WAR SYNDROME or PERSIAN GULF WAR or PERSIAN GULF SYNDROME. References of identified studies were searched for further studies. Specialist Gulf veterans' illnesses research websites (US Department of Defense Center for Deployment Health Research site and the Walter Reed Army Medical Center Gulf War database) and more-general Gulf websites were also searched for any additional references. Researchers who had expressed an interest in Gulf veterans' illness research were contacted for any non-published information. There was no restriction on the identification of studies in terms of publication status or language. This search strategy was first applied to data published up to the end of 1998 ($n=4156$) and then repeated to the end of May 2001 ($n=1231$).

Studies were included if they contained data on veterans who had been deployed to the Gulf War on military, medical or peace-keeping grounds (i.e. those involved in operations Desert Shield, Desert Storm, Granby or Desert Peace). Any study design was eligible for inclusion provided that an appropriate control or comparison group was included to compare the prevalence of psychiatric disorder.

The 5387 abstracts identified by the original search were screened by N.J.S. and the 2296 that remained eligible were examined by two members of the research team to decide whether they might meet our inclusion criteria. Printed copies of 409 papers were then obtained and examined by two members of the research team to confirm eligibility and extract data.

In our original search we also included studies that compared ill and well Gulf War veterans, but these were excluded from the review reported here. Studies were also excluded if they measured simulated exposures, if they measured non-health-related outcomes, or if the study population included inhabitants of the Persian Gulf states rather than deployed military, medical or peace-keeping personnel.

All identified papers that fulfilled the pre-stated inclusion criteria were categorised by health outcome. Forty-nine studies included data on psychiatric disorder, 29 of which reported on Gulf War veterans and an external comparison group of non-Gulf War veterans. We further restricted the studies to those with a limited range of outcomes concerned with psychiatric disorder (20 studies). The

outcomes we chose to include were as follows:

- PTSD diagnosed using a recognised standardised assessment;
- common mental disorder: depression or anxiety diagnosed using a recognised standardised assessment; or self-reported symptoms of depression recorded on a checklist;
- problems related to alcohol misuse.

We have chosen to use the term 'common mental disorder' (Goldberg & Huxley, 1992) to refer to the common symptoms of depression and anxiety that are seen in the community and reflect the use of assessments such as the General Health Questionnaire (GHQ; Goldberg & Williams, 1988) and the Symptom Checklist (and its derivatives) (Derogatis *et al.*, 1974; Derogatis, 1977; Derogatis & Spencer, 1982).

Data extraction

Data relating to the studies' main hypotheses and to methodological quality were extracted independently by two members of the research team. Information on the methodological quality of the individual studies included the response rate, the potential of selection bias in the sampling of the study participants, the potential bias in the measurement of outcomes, the availability of data on confounders, and any adjustment for such variables.

Statistical analysis

Summary odds ratio and risk ratios were calculated with a random-effects model using the inverse variance method. The degree of heterogeneity was assessed using the chi-squared test within a fixed-effects model. All analyses were performed using the METAN command (Bradburn *et al.*, 1998) in Stata version 6 (StataCorp, 1999). We chose this approach because of the inherent heterogeneity in the data. In particular, we were combining studies with a variety of outcome measures. A random-effects model assumes that the studies in a meta-analysis are sampled from a distribution of effect sizes, which are estimated from the data in the meta-analysis. In contrast, a fixed-effects model assumes that all the studies are sampled from a population with the same effect estimate.

We chose to perform analyses on dichotomous outcomes because the distribution of scores from continuous scales is

often difficult to establish from published articles, and this – together with the wide variety of scales that were used – can introduce difficulties in performing a quantitative synthesis. Using ratio measures to estimate association should be less sensitive to the different case definitions and measures used in the constituent studies.

RESULTS

The systematic review process is shown in Fig. 1. We identified 20 primary studies that investigated the association between deployment to the Gulf War and psychiatric disorder (Perconte *et al.*, 1993; Sutker *et al.*, 1993, 1994; Stretch *et al.*, 1996a,b; Iowa Persian Gulf Study Group, 1997; Pierce, 1997; Stuart & Halverson, 1997; Goss Gilroy Inc., 1998; Holmes *et al.*,

1998; Proctor *et al.*, 1998; Stuart & Bliese, 1998; Gray *et al.*, 1999; Ishoy *et al.*, 1999; Unwin *et al.*, 1999; Wolfe *et al.*, 1999; Bartone, 2000; Kang *et al.*, 2000; Steele, 2000; Cherry *et al.*, 2001). We excluded nine other studies that included data on psychiatric disorder in Gulf War veterans but did not meet our inclusion criteria: five repeated results already included, three did not include any of the psychiatric outcomes defined above, and one compared Gulf veterans with reported illness with a comparison sample (further details available from the authors upon request).

Table 1 summarises the studies we identified. All are best described as cross-sectional surveys. Some studies, for example those by Kang *et al.* (2000), Unwin *et al.* (1999) and Ishoy *et al.* (1999), resemble cohort studies, as the population was

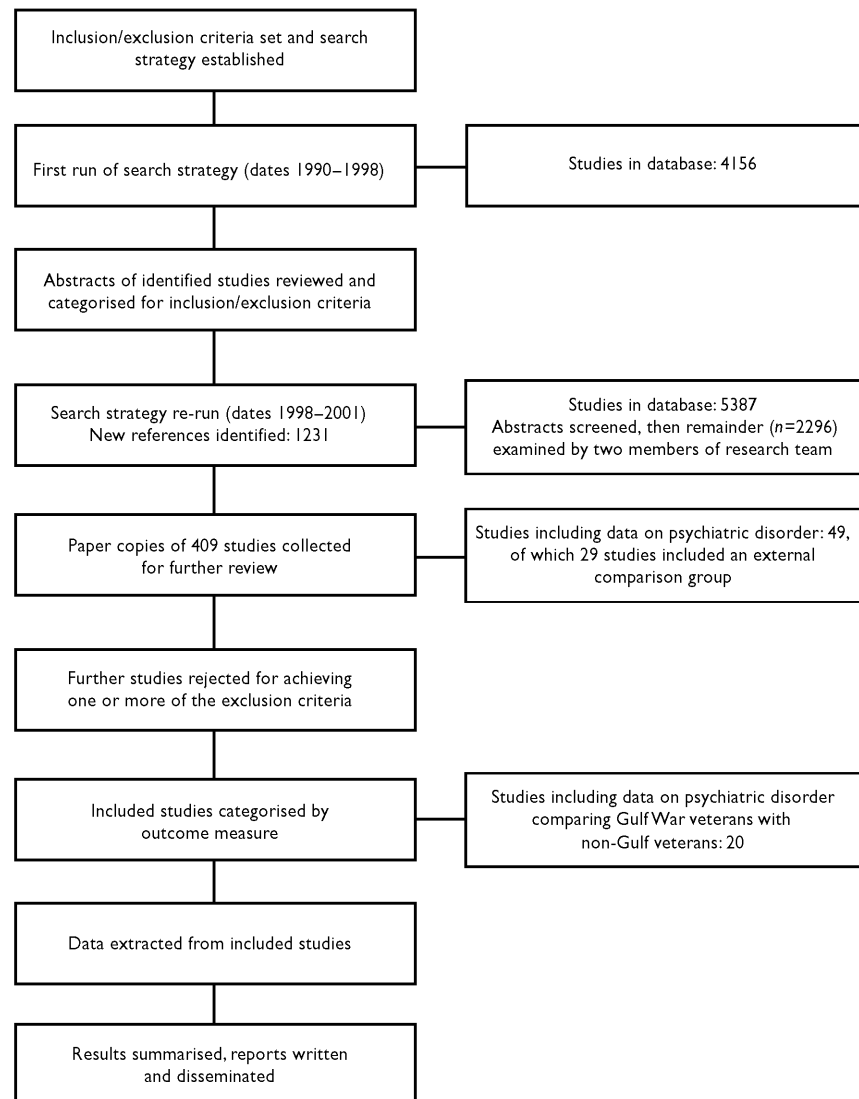


Fig. 1 Flowchart of the systematic review.

defined in terms of 'exposure' to the Gulf War. However, these studies had little or no information on health status before deployment and therefore share most of the methodological limitations of cross-sectional surveys.

Sampling

The sampling design of the studies varied. For example, Unwin *et al* (1999), Kang *et al* (2000), Goss Gilroy Inc. (1998), Ishoy *et al* (1999) and Cherry *et al* (2001) identified samples of service personnel from military databases. The Unwin *et al* and Cherry *et al* studies were of two independent samples drawn from the same UK military database. They employed stratified random sampling in order to frequency-match the characteristics of Gulf War veterans with those who were on active duty at the time but were not deployed to the Gulf. These comparison groups are referred to as non-Gulf veterans; the proportion actually deployed to areas other than the Gulf varied between studies. An alternative sampling strategy used by two studies, the Iowa Persian Gulf Study Group (1997) and Steele (2000), identified all military service personnel who had served during the period of the Gulf War and who lived in one US state (Iowa and Kansas, respectively). Within this standard survey design the investigators then compared those who had been deployed to the Gulf with those who had not. Pierce (1997) also used a military database but selected only women from the US Air Force to study.

There were also more *ad hoc* sampling procedures that did not use the large national databases. For example, Holmes *et al* (1998), Gray *et al* (1999) and Sutker *et al* (1993) compared Gulf War veterans and non-Gulf veterans within a selection of units. Some studies also chose a small number of military bases without any apparent justification for inclusion (Proctor *et al*, 1998; Wolfe *et al*, 1999).

Response rates

Response rates also varied considerably between studies (Table 1). Of most importance is that the response rate of the Gulf War veterans was higher than that of the non-Gulf veterans in studies that reported the response rates separately. This could introduce a biased comparison. For example, Unwin *et al* (1999) had a 70% response rate in the Gulf War veterans

and a 63% response rate in the non-Gulf War veterans sample. Goss Gilroy Inc. (1998) in the Canadian study reported response rates of 73% for Gulf War veterans and 60% for non-Gulf War veterans.

Measurement

Most of the studies took place after there had been considerable publicity about illness in Gulf War veterans. However, four studies included here reported findings based upon surveys carried out within about a year of the end of the Gulf War: these studies were by Sutker *et al* (1993, 1994), Holmes *et al* (1998) and Stuart & Halverson (1997). All reported a significant excess of psychopathological disorder within the Gulf War veterans.

Many of the studies used the Mississippi scale (Keane *et al*, 1988) or modified versions thereof to assess symptoms of PTSD; this is a self-administered scale and it is generally assumed to be less valid than some of the more detailed questionnaires. Some studies used their own method for assessing PTSD based upon questions modelled on the DSM-III-R (American Psychiatric Association, 1987) criteria. A few studies used structured interviews administered by clinicians (Sutker *et al*, 1994; Proctor *et al*, 1998; Wolfe *et al*, 1999), but these assessments would have had the potential disadvantage of introducing possible observer bias, as the interviewers would not have been masked to the participants' deployment status.

We identified 17 studies that included data on common mental disorders. The self-administered questionnaire used most frequently to assess common mental disorder, in eight studies, was the Hopkins Symptom Checklist or Brief Symptom Inventory (Derogatis *et al*, 1974; Derogatis, 1977; Derogatis & Spencer, 1982; Derogatis & Melisaratos, 1983). This scale was reported either as a continuous outcome or used to define a 'case' of common mental disorder. The other studies used a variety of methods to assess common mental disorder, from self-reported symptoms of depression (Proctor *et al*, 1998; Gray *et al*, 1999; Ishoy *et al*, 1999; Kang *et al*, 2000; Steele, 2000), other self-administered scales such as the GHQ (Unwin *et al*, 1999), to lengthy clinician-administered structured interviews (Wolfe *et al*, 1999).

Confounding

There was considerable variation in the extent to which the authors attempted to adjust for confounders. Many of the studies that selected from the military databases used a stratified sampling procedure and frequency-matched the non-Gulf veterans on some characteristics in order to adjust for confounding. Some studies included these variables in a multivariate model when analysing their results, which was probably necessary given the differential response rate between the Gulf War veterans and non-Gulf veterans. The most thorough adjustments were carried out by Unwin *et al* (1999). In particular, only Unwin *et al* and Stuart & Bliese (1998) adjusted for marital status. This is likely to be an important confounding variable, as single people usually have higher rates of common mental disorder and were more likely to be deployed to the Gulf War – although not in Unwin *et al*'s study, possibly because the UK military have fewer members who are never deployed on active service. Unwin *et al* (1999) found that the odds ratio for being a case on the GHQ changed only from 2.0 to 2.1 after adjustment, indicating that there was little evidence of confounding by the variables identified in that study. Results similar to these were obtained using PTSD as the outcome.

Meta-analysis

Post-traumatic stress disorder

It was possible to conduct a meta-analysis of 9 of the 11 studies that reported dichotomous outcomes for PTSD. We were unable to use the data from Goss Gilroy Inc. (1998) and Bartone (2000). The results are summarised in Fig. 2. The overall summary estimate using a random-effects model was an odds ratio of 3.17 (95% CI 2.16–4.65), indicating an increased risk in Gulf War veterans. There was significant heterogeneity ($\chi^2=29.4$, d.f.=8, $P<0.0001$). In particular, the two large studies by Unwin *et al* and Gray *et al* differed: the former found an OR of 3.5 and the latter on OR of 1.8. The summary estimate for the risk ratios was 2.9 (95% CI 2.0–4.2).

Common mental disorder

We were able to perform a meta-analysis on 11 of the studies that reported on the prevalence of common mental disorder (Fig. 3). Two studies used the same

Table 1 Studies reporting the prevalence of psychiatric disorder in Gulf War veterans (GWV) and veterans not deployed to the Gulf (NGV). All studies are of US servicemen unless otherwise stated

Reference	Study design	Sample	Study period	Main outcomes	Main results	Selection or response bias	Confounding
Samples from national military databases							
Pierce (1997)	Longitudinal study comparing GWV (n=153) with those deployed elsewhere (n=331)	Stratified random sample, female US Air Force GWV and 'other deployed' veterans. Sampling frame from US DoD manpower data centre. Stratified by active duty/reserve, parent/non-parent	T ₁ was 2 years after war and T ₂ 4 years	Sub-scales for depression, anxiety and somatisation based on Hopkins Symptom Checklist. Depression item from RAND Questionnaire	Common mental disorder	Started with 638 women, traced 525, of whom 92% and 87% responded at T ₁ and T ₂ respectively	Controlled for age using analysis of variance. Possibly frequency-matched for service status and parent
				Mean scores on depression item (no s.d. given): T ₁ : NGV 1.84, GWV, 2.02 T ₂ : NGV 1.66, GWV 1.87			
				Mississippi scale	PTSD		
				GWV 24% NGV 15%			
				Unclear whether at T ₁ or at T ₂			
Goss Gilroy Inc. (1998)	Cross-sectional survey: postal	All Canadian GWV (n=3113) identified from the Canadian Department of National Defense and a 'deployed elsewhere' control group matched for gender, age, reserve/regular (n=3439)	Survey June–Dec. 1997	Minor and major depression (PRIME-MD) Alcohol misuse (measure not specified)	Common mental disorder	Response rate: GWV 73.0% NGV 60.3%	In addition to the matching, the ORs were adjusted for other confounders that were statistically significantly associated with outcome
				PTSD based on symptom report (PCL-M)			
				PTSD			
				PTSD OR=2.69 (95% CI 1.59–4.26) adjusted for income			
Stuart & Bliese (1998)	Cross-sectional survey: postal	Randomly selected units from all US Army National Guard and Reserves taken from the DoD manpower data centre; NGV were deployed to USA or Germany GWV n=991 NGV n=279	Jan. and Feb. 1993	Brief Symptom Inventory, Cases defined on the Global Severity Index	Common mental disorder	Return rate was 31%	Multiple regression was used to control for rank, gender, age, marital status, ethnicity, current life stress scale, NGV group from USA omitted because of error in table

Table 1 (continued)

Reference	Study design	Sample	Study period	Main outcomes	Main results	Selection or response bias	Confounding
Ishoy <i>et al</i> (1999)	Cross-sectional survey	All Danish military GWV and NGO aid workers compared with NGV group frequency-matched on gender, age and 'profession' using a Danish military database GWV <i>n</i> =686 NGV <i>n</i> =231	Feb. 1997–Jan. 1998	Common mental disorder Self-reported symptom of depression or sadness during previous 12 months that started during or after Gulf War	Prevalence: GWV 22.6% NGV 10.4%	Response rate: GWV 83.6% NGV 57.8%	Controls matched for gender, age and profession. Non-response meant GWV more likely to be older males
Unwin <i>et al</i> (1999)	Cross-sectional survey: postal	Stratified random sample drawn from UK military database: GWV <i>n</i> =3284 Bosnian veterans <i>n</i> =1815 Era veterans <i>n</i> =2408 Bosnia and era samples frequency-matched for age, service, service status, gender, rank and fitness level Different sample from Cherry <i>et al</i> (2001)	Aug. 1997–II Nov. 1998	Common mental disorder Case was ≥3 on GHQ-12 (Goldberg & Williams, 1988) Era 24.0% GWV v. Bosnia OR=1.6 (95% CI 1.4–1.8) GWV v. era, OR=2.1 (95% CI 1.9–2.4) Adjusted ORs reported PTSD	Prevalence: GWV 39.2% Bosnia 26.3% Era 24.0% GWV v. Bosnia OR=1.6 (95% CI 1.4–1.8) GWV v. era, OR=2.1 (95% CI 1.9–2.4) Adjusted ORs reported	Response rate: GWV 70.4% Bosnia 61.9% Era 62.9% Responders were significantly older and more likely to be in service, but did not differ on SF-36	Potential confounders (age, marital status, rank, education, employment, still serving or discharged, smoking and alcohol consumption) are adjusted for using logistic regression. The samples were frequency-matched by age, service, service status, gender, rank and fitness. Analyses restricted to men only
Kang <i>et al</i> (2000)	Cross-sectional survey	Stratified random sample taken from the DoD manpower data centre GWV <i>n</i> =11 441 NGV <i>n</i> =9476 Frequency-matched on gender and service status	Not mentioned	Common mental disorder Self-reported symptom of 'depression' Rate difference 14% (95% CI 13.9–14.1)	Prevalence: GWV 36% NGV 22% Rate difference 14% (95% CI 13.9–14.1)	Response rate: GWV 75% NGV 64% Non-responders were more likely	No adjustment made for potential confounders except for the matching variables, gender and service status

Table 1 (continued)

Reference	Study design	Sample	Study period	Main outcome	Main results	Selection or response bias	Confounding
Cherry <i>et al</i> (2001)	Cross-sectional survey	Stratified random sample from the UK Ministry of Defence database GWV (<i>n</i> =8014) and NGV (<i>n</i> =3900) frequency-matched by gender, age, service and rank different sample from Unwin <i>et al</i> (1999)	Dec. 1997–Sept. 1999	Common mental disorder Self-reported symptom 'Feeling unhappy and depressed': visual analogue scale	Mean score: GWV 6.3 NGV 3.7 (from figure)	Response rate: GWV 84.3% NGV 82.1%	No adjustment apart from matching variables: gender, age, service and rank
Representative samples restricted to a single geographical area							
lowa Persian Gulf Study Group (1997)	Cross-sectional telephone survey	DoD manpower data centre used to create a stratified random sample from 28 968 military personnel from Iowa: GWV <i>n</i> =1896 NGV <i>n</i> =1799	Sept. 1995–May 1996	Common mental disorder Brief Symptom Inventory – case definition not stated CAGE questionnaire for alcohol misuse – case definition not stated	Prevalence difference in depression GWV v. NGV, 6.0 (95% CI 4.0–7.9) Prevalence of depression (estimated from tables): GWV=9.0% NGV=4.6% Prevalence of alcohol misuse (estimated from tables): 17.4% v. 12.6%, <i>P</i> =0.02	Response rate: GWV 78.3% NGV 73%	Adjusted for stratification variables: military service (regular/National Guard) age, gender, race, branch of service and rank
PTSD							
Steele (2000)	Cross-sectional telephone survey	GWV <i>n</i> =1435, NGV <i>n</i> =409 DoD manpower data centre used to create a stratified random sample from 16 566 military personnel from Kansas. Stratified by reservist and gender	Feb.–Aug. 1998	Common mental disorder Self-reported 'Feeling down or depressed'	Prevalence (estimated from tables): GWV 1.9% NGV 0.8% Prevalence: GWV 23% NGV 9% OR = 2.99 (95% CI 2.07–4.31)	Response rate: 65% (3138 eligible) Gulf and female veterans were more likely to respond	Odds ratio adjusted for gender, age, income and education level

Table 1 (continued)

Reference	Study design	Sample	Study period	Main outcomes	Main results	Selection or response bias	Confounding
Other sampling strategies							
Perconte <i>et al</i> (1993)	Cross-sectional survey	Selected from 'various' reserve units in western Pennsylvania: NGV <i>n</i> =126 GWV <i>n</i> =439 EDV <i>n</i> =26	Not provided	Cases defined on Global Severity Index (score ≥0.7) or SCL-90-R; Beck Depression Inventory case score ≥ 10 Administered after psycho-educational presentations and discussion	Common mental disorder Prevalence on GSI: GWV 27.9% EDV 11.5% BGV 15.1% on BDI: GWV 26.9% EDV 23.1% NGV 16.7%	Response rate: 95% (approximate)	No adjusted results presented
PTSD							
Mississippi scale							
Prevalence: GWV 15.5% NGV 4.0% EDV 3.9%							
Sutker <i>et al</i> (1993)	Cross-sectional survey	Participants were drawn from 5 National Guard and Army Reserve units in Louisiana: 215 troops deployed to the Gulf and 60 troops from the same units who were activated but not deployed overseas	Troops were assessed 4–10 months after completion of Desert Storm	Beck Depression Inventory; anxiety: State-Trait Anxiety Inventory; Health Symptom Checklist (includes 9 items from Hopkins Symptom Checklist)	Common mental disorder The GWV who experienced 'high war-zone stress' scored significantly more highly on the BDI and STAI anxiety measures	Response rate: GWV 70.3% No response rate given for NGV	All personnel had returned to the USA and not sought mental health treatment services. GWV were more likely to be younger and higher military rank, although correlational analyses found no significant difference between these measures and scores on mood measures
Sutker <i>et al</i> (1994)	Cross-sectional survey	60 Army reservists assigned grave registration duties GWV (<i>n</i> =40) NGV (<i>n</i> =20) 'activated but remained stateside'	12 months after return from Desert Storm	Depressive disorder: NOS; SCID-P	Common mental disorder Higher prevalence in GWV (13%). No case found in NGV	The 60 respondents were selected from 124 in a survey of unspecified design. Response rate uncertain	All troops with previous combat experience were GWV GWV older than NGV No adjustment for confounding

Table 1 (continued)

Reference	Study design	Sample	Study period	Main outcome	Main results	Selection or response bias	Confounding
<i>PTSD</i>							
Stretch et al (1996a)	Cross-sectional survey: postal	Various units from Hawaii and Pennsylvania who either deployed to the Gulf or did not deploy GWV n=1524 NGV n=2512	1993	Defined using SCID-P	Higher prevalence in GWV. No cases in NGV		
				Common mental disorder		Response rate: 31%	Hierarchical multiple regression used to adjust for age, rank, service branch, race, education, drinking and smoking
Stretch et al (1996b)	Cross-sectional survey: postal	Various units from Hawaii and Pennsylvania who either deployed to the Gulf or did not deploy GWV n=1524 NGV n=2512	1993	Walter Reed Army Institute of Research PTSD algorithm (9 items from IES) and 8 items from BSI	Prevalence: GWV 8.6% NGV 1.6%	Response rate: 31%	No adjustment was made for potential confounders. Current life stresses were found to be strongly related to PTSD symptoms
				Brief Symptom Inventory	GWV significantly higher on all sub-scales (and after adjustment)		
<i>PTSD</i>							
Stuart & Halverson (1997)	Cross-sectional survey	No details provided of sampling strategy GWV n=2180 male, n=182 female Bosnia veterans n=1254 male, n=184 female, serving in Bosnia May-July 1996	GWV were studied from Nov. 1990-Mar. 1991; NGV were sampled at 2 time points; May-Sep. 1991 and Jan.-Mar. 1996	Common mental disorder	GSI mean (s.d.): GWV, male 0.7 (0.7), female 0.8 (0.7) Bosnia veterans, male 0.6 (0.6), female 0.7 (0.6)	Response rate: unknown	No apparent adjustment for confounders
				Brief Symptom Inventory			
Holmes et al (1998)	Cross-section survey: postal	All members of an Air National Guard Unit GWV n=296 NGV n=210	11 months after end of hostilities	Common mental disorder	Case defined as score ≥70 on Global Severity Index of SCL-90-R	Response rate: GWV 57.2% NGV 42.3%	Used logistic regression but insufficient details to decide what adjustments were made
					Prevalence: GWV 11.5% NGV 7.3% Significant increase in mean scores on GSI in GWV		

Table 1 (continued)

Reference	Study design	Sample	Study period	Main outcome	Main results	Selection or response bias	Confounding
PTSD							
Proctor <i>et al</i> (1998)	Cross-sectional survey	Stratified random samples of GWV from Fort Devens, New England (<i>n</i> =186) and New Orleans (<i>n</i> =66), and an NGV comparison group from an air ambulance unit (<i>n</i> =48)	Spring 1994 to autumn 1996	Mississippi scale (case ≥ 90)	Prevalence: GWV 6.8% NGV 1.7%		
				Self-reported 'frequent periods of feeling depressed'	Prevalence: New England 22.6% New Orleans 5.8% NGV 1.6% OR=6.0 for New England, 3.9 for New Orleans	Response rate: New England 62% New Orleans 38% NGV 51% Participants were recruited after taking part in a previous study whose response rate was 78% NGV not very comparable	Prevalence estimates account for stratification. Odds ratios adjusted for age, gender and education with logistic regression
PTSD							
Wolfe <i>et al</i> (1999)	Cross-sectional survey	Stratified random samples of GWV from New England (Fort Devens) (<i>n</i> =148) and New Orleans (<i>n</i> =56) and an NGV comparison group from an ambulance unit (<i>n</i> =48) (same study as Proctor <i>et al</i> (1998))	1994–1996	CAPS Mississippi scale for Desert Storm war zone personnel. CAPS possibly subject to observer bias	Prevalence, CAPS: New England 5% (8/148) New Orleans 8% (4/58) NGV 0% Prevalence, Mississippi: New England 8.1% New Orleans 7.6% NGV not stated		
				SCID used to define DSM-III-R major depressive disorder CAPS and SCID administered by trained clinicians Blindness to exposure status not mentioned	Common mental disorder	Major depressive disorder: New England GWV 6.6% New Orleans GWV 4.5% NGV 0%	Response rate: GWV 30–42% NGV 51% of unit GWV non-responders tended to be younger and unmarried. The GWV groups were quite different from the NGV group

Table 1 (continued)

Reference	Study design	Sample	Study period	Main outcome	Main results	Selection or response bias	Confounding
Gray et al (1999)	Cross-sectional survey	Active duty Seabees (naval construction workers) who remained in service after Desert Storm/Shield and were serving in one of two large Seabee centres. Selected using the DoD manpower database GWV n=527 NGV n=970	Sept. 1994–June 1995	Five dimensions of Hopkins Symptom Checklist Self-reported symptom of depression	Common mental disorder GWV had statistically significantly increased scores on all 5 dimensions of the Hopkins Symptom Checklist Prevalence of self-reported depression: GWV 6.8% NGV 2.8% OR=2.6 (95% CI 1.5–4.4) PTSD	Estimated overall response rate: 53%	GWVs younger, more often male and less educated No adjustments for confounders
Bartone (2000)	Cross-sectional survey	Six army reserve medical units GWV n=389 NGV n=381 (236 to USA, 145 to Germany)	4–6 months after the end of the Gulf War	PTSD screen (items from DSM-IV) Impact of Events Scale	PTSD Prevalence: GWV – 15.2% NGV 9.0% OR 1.8 (95% CI 1.3–2.5)	Response rate: approximately 50%	Statistically significant after adjustment for number of stressful life events and 'hardness'

BDI, Beck Depression Inventory (Beck et al, 1961); CAPS, Clinician Administered PTSD Scale (Blake et al, 1990); DoD, Department of Defense; EDV, Europe-deployed veteran; GHQ-12, 12-item General Health Questionnaire (Goldberg & Williams, 1998); GSI, Global Severity Index (Derogatis & Spencer, 1982); Derogatis & Melisaratos, 1983; GWV, Gulf War veteran; IES, Impact of Event Scale; NGO, non-governmental organisation; NGV, non-Gulf veteran; NOS, not otherwise specified; PCL-M, Post-traumatic stress disorder Checklist – Military (Weathers et al, 1993); PRIME-MD, Primary Care Evaluation of Mental Disorders (Spitzer et al, 1974); PTSD, post-traumatic stress disorder; SCID, Structured Clinical Interview for DSM-III-R (Spitzer et al, 1990); SCL-90-R, Symptom Checklist (Derogatis, 1977); SF-36, 36-item Short Form Health Survey; STAI, State-Trait Anxiety Inventory (Spielberger et al, 1970).
1. This study included two non-Gulf veteran groups: the first, the era group, comprised personnel who were serving in the military at the time, and the second comprised personnel who were deployed in Bosnia (many of whom were not serving at the time of the Gulf War).

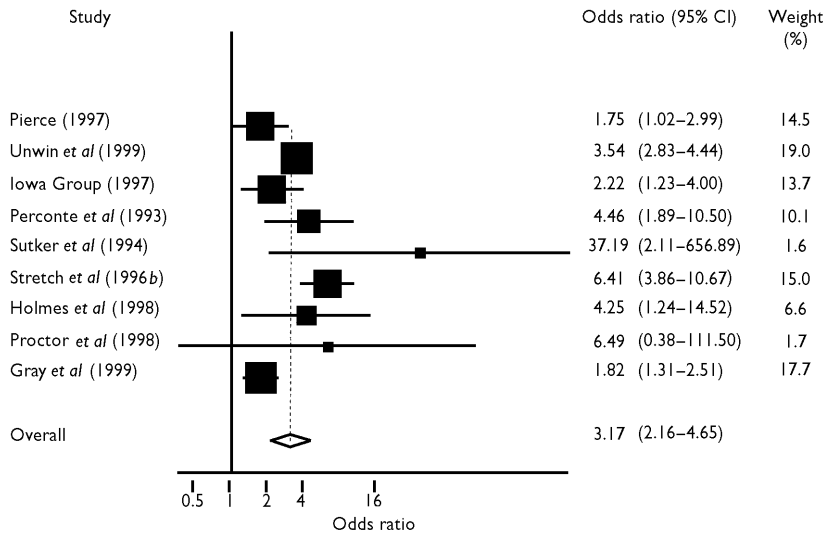


Fig. 2 Forest plot of the results for post-traumatic stress disorder. Odds ratios by deployment. Values above 1 indicate increased risk in Gulf War veterans

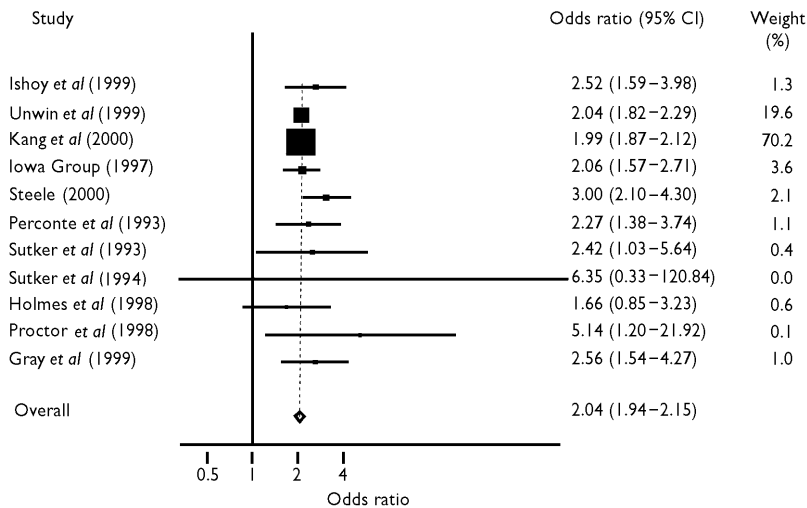


Fig. 3 Forest plot of the results for common mental disorder. Odds ratios by deployment. Values above 1 indicate increased risk in Gulf War veterans.

sample, but one (Wolfe *et al*, 1999) reported results from the Structured Clinical Interview for DSM-III-R (Spitzer *et al*, 1990) and the other (Proctor *et al*, 1998) presented results from self-reported symptoms of depression. The summary estimate was an odds ratio of 2.04 (95% CI 1.94–2.15), irrespective of whether the data from either of these studies were excluded, indicating an increased risk of common mental disorder in the deployed service personnel. Despite the variation between studies in the outcome used, there was no statistical evidence to support heterogeneity in this sample using odds ratios (heterogeneity test $\chi^2=9.39$, d.f.=10, $P=0.5$). The summary estimate for risk ratio was 1.8 (95% CI 1.6–2.0).

It should be noted that the studies by Kang *et al* (2000) and Unwin *et al* (1999) accounted for 90% of the variance weights in the meta-analysis. The other studies therefore had little influence on the summary estimate.

A funnel plot of the standard error of the estimate against the size of effect suggested that there were fewer small non-significant findings than would be expected. This would not have had much influence on the findings, given the presence of a number of large studies.

Alcohol misuse

There was little evidence concerning alcohol misuse or dependence. Goss Gilroy

Inc. (1998) stated that there was no statistically significant association between alcohol misuse and deployment. The Iowa study (Iowa Persian Gulf Study Group, 1997) reported an increased prevalence of alcohol misuse measured by the CAGE questionnaire (Ewing, 1984).

DISCUSSION

The results of our systematic review and meta-analysis indicate an increased prevalence of PTSD and common mental disorder in service personnel who had been deployed to the Persian Gulf War. The size of this effect was somewhat larger for PTSD, with an OR of 3.2 (95% CI 2.2–4.7) compared with 2.0 (95% CI 1.9–2.1) for common mental disorder.

Publication bias

We adopted a thorough search strategy but – as in all systematic reviews – may have failed to identify some studies. We are also aware that other studies on this topic are in progress and have yet to report their findings. It is difficult to assess the effect of any publication or citation bias in our data, given the small number of studies that reported data in a form permitting meta-analysis. A funnel plot for the ‘common mental disorder’ outcome suggested that there was an underrepresentation of small studies finding no association between deployment to the Gulf and disorder. However, these small studies would not have had a major impact on the summary odds ratio, despite suggesting that it might be a slight overestimate. The summary estimate was dominated by the two large studies.

Sample selection

A critical part of these designs is the comparability of the deployed and non-deployed troops. Some of the studies used military databases and took care to ensure that their sample was representative of both Gulf War veterans and the comparison group. It is likely that the characteristics of troops selected for deployment systematically differed from those of other active service personnel who were not deployed. This could be less marked for the UK military service, in which almost everyone is likely to be deployed on active duty. Potential confounding factors include gender, fitness level and marital status, along with

other aspects (such as propensity for risk-taking) that are more difficult to measure. It is also likely that within individual units, the reasons for choosing people for deployment would lead to a greater selection bias than in studies sampling from national databases, in which whole units would have been selected.

It is difficult to be sure about the effect of selection on the results reported here. Some authors have suggested a 'healthy warrior' effect, that the deployed have better underlying health. On the other hand, single people, who are more likely to have been deployed (at least in some studies; Proctor *et al*, 1998), tend to have poorer mental health (Kessler *et al*, 1994; Jenkins *et al*, 1997). None of the studies had any independent information about the mental health of participants before the Gulf War and so were not able to take any account of this factor.

Non-response bias

The studies that reported response rates according to deployment status all found that the Gulf War veterans had a higher response rate. It is likely that the publicity surrounding illnesses in Gulf War veterans increased the relevance of a questionnaire about health effects to respondents who had been deployed to the Gulf. This differential response rate could introduce a systematic bias.

Some studies have reported that non-responders tended to have poorer mental health than those who responded (Williams & Macdonald, 1986) although Unwin *et al* (1999) in a more intensive follow-up of non-respondents did not find a statistically significant increased risk of common mental disorder. Kang *et al* (2000) also compared those who responded to the later mailings with those who returned the first mailshot. They did not find that the later respondents had poorer self-rated general health. On balance, it is unlikely that the differential response rate seen in these studies could have explained such a large association as that reported.

Outcome measurements

The majority of studies relied on self-reported symptoms to assess the prevalence of psychiatric disorder. Some of the studies used well-recognised and validated measures of psychiatric disorder, but others (including some of the larger studies) reported results from a single question asking about

depressive symptoms. Despite this variation in measurement methods, there was little evidence of heterogeneity in the estimates for common mental disorder. Studies that used the longer semi-structured interviews might have introduced observer bias, given the difficulty in 'blinding' the interviewers. In contrast, there was evidence of heterogeneity for the PTSD estimates. In particular, Unwin *et al* (1999) reported a larger effect than did Gray *et al* (1999), although both reported a significant increase in prevalence in the Gulf War veterans. Gray *et al* restricted their sample to naval construction workers, so the different result might merely have reflected the different experiences of this group of service personnel. It should also be noted that the Unwin *et al* study used a UK military cohort in which almost all the non-Gulf War veterans comparison group would have been deployed on active service at one time or another.

Five of the 20 studies were carried out within 12 months of the end of the war and at a time when publicity concerning illness in Gulf War veterans was minimal. All these studies reported a statistically significant increase in psychopathological disorders in Gulf War veterans. These early studies tended to be less robust from a methodological point of view than the later ones: the samples were less representative, response rates were lower and the studies smaller in size. In contrast, the later and often more robust studies could have been subject to a reporting bias following publicity about illnesses in Gulf War veterans. In conclusion, it appears unlikely that a reporting bias could have led to the findings reported in the constituent studies.

Illnesses in Gulf War veterans

We found that veterans deployed to the Persian Gulf War reported more PTSD and more symptoms of common mental disorder than did service personnel who had not been deployed to this war. Increased rates of PTSD have often been reported after conflicts and can be attributed to the increased likelihood of psychologically traumatising events during wartime. The increased rates of other psychiatric symptoms might just be a consequence of the same process. There is evidence that psychologically traumatic events also lead to an increase in other psychological symptoms, particularly anxiety, in addition to the symptoms more specifically associated with the syndrome of

PTSD. An increased rate of psychiatric disorders would therefore be expected in Gulf War veterans, although this does not diminish the importance of this morbidity in affecting veterans many years after returning from the conflict.

What is less clear is how these findings relate to the issue of Gulf War illnesses. Gulf War veterans have reported a wide variety of symptoms, aside from psychiatric symptoms. Unwin *et al* (1999), in the UK study, reported an increased prevalence of a whole range of symptoms after having adjusted for the increased prevalence of common mental disorder in the Gulf War veterans. This supports the view that some other factors must be contributing to illnesses in these veterans, in addition to any increase in psychiatric disorder.

Psychiatric disorder is common, disabling and burdensome. It is an important source of disability after war, yet this is often inadequately recognised and acknowledged. Developing more-effective means of preventing and treating psychiatric disorder in service personnel is an important priority for future research.

ACKNOWLEDGEMENTS

We thank Simon Wessely and Matthew Hotopf for comments on an earlier draft of the manuscript. We thank the Department of Information Services at University of Wales College of Medicine for assistance in obtaining references for this review.

REFERENCES

- American Psychiatric Association (1987)** *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn, revised) (DSM-III-R). Washington, DC: APA.
- Bartone, P. T. (2000)** Hardiness as a resiliency factor for United States forces in the Gulf War. In *Posttraumatic Stress Intervention: Challenges, Issues and Perspectives* (ed. D. Paton), pp. 115–133. Springfield, IL: Charles C. Thomas.
- Beck, A. T., Ward, C. H., Mendelsohn, M., et al (1961)** An inventory for measuring depression. *Archives of General Psychiatry*, **4**, 561–571.
- Blake, D., Weathers, F., Nagy, L., et al (1990)** *Clinician Administered PTSD Scale*. Boston, MA: National Center for PTSD, Behavioral Sciences Division.
- Bradburn, M. J., Deekes, J. J. & Altman, D. G. (1998)** METAN – an alternative meta-analysis command. *Stata Technical Bulletin*, **44**, 4–15.
- Cherry, N., Creed, F., Silman, A., et al (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.
- Derogatis, L. R. (1977)** *The SCL-90-R Manual I: Scoring, Administration and Procedures for the SCL-90*. Baltimore, MD: Johns Hopkins University School of Medicine.

— & Melisaratos, N. (1983) The Brief Symptom Inventory: an introductory report. *Psychological Medicine*, **13**, 595–605.

— & Spencer, P. M. (1982) *The Brief Symptom Inventory: Administration, Scoring and Procedures Manual*. Baltimore, MD: Johns Hopkins University School of Medicine, Clinical Psychometrics Research Unit.

—, Lipman, R. S., Rickels, K., et al (1974) The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behavioural Science*, **19**, 1–15.

Ewing, J. A. (1984) Detecting alcoholism: the CAGE questionnaire. *JAMA*, **252**, 1905–1907.

Goldberg, D. & Huxley, P. (1992) *Common Mental Disorders: A Biopsychosocial Approach*. London: Routledge.

— & Williams, P. (1988) *The User's Guide to the General Health Questionnaire*. Windsor: NFER–Nelson.

Goss Gilroy Inc. (1998) *Health Study of Canadian Forces Personnel Involved in the 1991 Conflict in the Persian Gulf* (vol. 1). Ottawa: Goss Gilroy Inc. http://www.dnd.ca/menu/press/Reports/Health/health_study_e_voll_TOC.htm

Gray, G. C., Kaiser, K. S., Hawksworth, A. W., et al (1999) Increased postwar symptoms and psychological morbidity among US Navy Gulf War veterans. *American Journal of Tropical Medicine and Hygiene*, **60**, 758–766.

Holmes, D. T., Tariot, P. N. & Cox, C. (1998) Preliminary evidence of psychological distress among reservists in the Persian Gulf War. *Journal of Nervous and Mental Disease*, **186**, 166–173.

Horowitz, M., Wilner, N. & Alvarez, W. (1979) Impact of Event Scale: a measure of subjective stress. *Psychosomatic Medicine*, **41**, 209–218.

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

Ishoy, T., Suadcani, P., Guldager, B., et al (1999) State of health after deployment in the Persian Gulf – the Danish Gulf War Study. *Danish Medical Bulletin*, **46**, 416–419.

Jenkins, R., Lewis, G., Bebbington, P., et al (1997) The National Psychiatric Morbidity Surveys of Great Britain: initial findings from the Household Survey. *Psychological Medicine*, **27**, 775–790.

Kang, H. K., Mahan, C. M., Lee, K. Y., et al (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.

Keane, T. M., Caddell, J. M. & Taylor, K. L. (1988) Mississippi Scale for Combat-Related Posttraumatic Stress Disorder: three studies in reliability and validity. *Journal of Consulting and Clinical Psychology*, **56**, 85–90.

Kessler, R. C., McGonagle, K. A., Zhao, S., et al (1994) Lifetime and 12-month prevalence of DSM–III–R psychiatric disorders in the United States. *Archives of General Psychiatry*, **51**, 8–19.

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

Perconte, S. T., Wilson, A. T., Pontius, E. B., et al (1993) Psychological and war stress symptoms among deployed and non-deployed reservists following the Persian Gulf War. *Military Medicine*, **158**, 516–521.

Pierce, P. F. (1997) Physical and emotional health of Gulf War veteran women. *Aviation, Space and Environmental Medicine*, **68**, 317–321.

CLINICAL IMPLICATIONS

- Veterans of the Gulf War report more post-traumatic stress disorder and more depression and anxiety than do war veterans not deployed to that conflict.
- Service in a war zone leads to an increase in symptoms many years afterwards.
- The presence of psychiatric symptoms probably does not explain the increased prevalence of other somatic symptoms reported by Gulf War veterans.

LIMITATIONS

- Most of the large studies were conducted after public concern about illnesses in Gulf War veterans had been voiced.
- The studies relied upon self-reported information about psychiatric symptoms.
- Some of the larger studies used non-standard methods of assessing psychiatric symptoms.

NICOLA J. STIMPSON, PhD, HOLLIE V. THOMAS, DPhil, ALISON L. WEIGHTMAN, PhD, FRANK DUNSTAN, PhD, University of Wales College of Medicine, Cardiff; GLYN LEWIS, FRCPsych, University of Bristol, UK

Correspondence: Professor Glyn Lewis, Cotham House, Cotham Hill, Bristol BS6 6JL, UK. E-mail: glyn.lewis@bristol.ac.uk

(First received 23 May 2002, final revision 18 October 2002, accepted 30 October 2002)

Proctor, S. P., White, R. F., Wolfe, J., et al (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.

Spielberger, C. D., Gorsuch, R. L. & Lushene, R. E. (1970) *STA: Manual for the State–Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Spitzer, R. L., Williams, J. B. W., Gibbon, M., et al (1990) *User's Guide for the Structured Clinical Interview for DSM–III–R: SCID*. Washington, DC: American Psychiatric Association.

—, —, Kroenke, K., et al (1994) Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME–MD 1000 study. *JAMA*, **272**, 1749–1756.

StataCorp (1999) *Stata Statistical Software*. College Station, TX: Stata Corporation.

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.

Stretch, R. H., Bliese, P. D., Marlowe, D. H., et al (1996a) Psychological health of Gulf War-era military personnel. *Military Medicine*, **161**, 257–261.

—, Marlowe, D. H., Wright, K. M., et al (1996b) Post-traumatic stress disorder symptoms among Gulf War veterans. *Military Medicine*, **161**, 407–410.

Stuart, J. A. & Bliese, P. D. (1998) The long-term effects of operation desert storm on the psychological distress of US army reserve and national guard veterans. *Journal of Applied Social Psychology*, **28**, 1–22.

— & Halverson, R. R. (1997) The psychological status of US Army soldiers during recent military operations. *Military Medicine*, **162**, 737–743.

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

—, —, —, et al (1994) Psychopathology in war-zone deployed and non-deployed operation desert storm troops assigned registration duties. *Journal of Abnormal Psychology*, **103**, 383–390.

Unwin, C., Blatchley, N., Coker, W., et al (1999) Health of UK servicemen who served in Persian Gulf War. *Lancet*, **353**, 169–178.

Weathers, F., Litz, B. T., Heran, D. S., et al (1993) *The PTSD Checklist (PCL): Reliability, Validity, and Diagnostic Utility*. San Antonio, TX: International Society of Traumatic Stress Studies.

Williams, P. & Macdonald, A. (1986) The effect of non-response bias on the results of two-stage screening surveys of psychiatric disorder. *Social Psychiatry*, **21**, 182–186.

Wolfe, J., Proctor, S. P., Erickson, D. J., et al (1999) Relationship of psychiatric status to Gulf War veterans' health problems. *Psychosomatic Medicine*, **61**, 532–540.