Gulf war illness—better, worse, or just the same? A cohort study
M Hotopf, Anthony S David, Lisa Hull, Vasilis Nikalaou, Catherine Unwin, Simon Wessely

Abstract

Objectives Firstly, to describe changes in the health of Gulf war veterans studied in a previous occupational cohort study and to compare outcome with comparable non-deployed military personnel. Secondly, to determine whether differences in prevalence between Gulf veterans and controls at follow up can be explained by greater persistence or greater incidence of disorders.

Design Occupational cohort study in the form of a postal survey.

Participants Military personnel who served in the 1991 Persian Gulf war; personnel who served on peacekeeping duties to Bosnia; military personnel who were deployed elsewhere (“Era” controls). All participants had responded to a previous survey.

Setting United Kingdom.

Main outcome measures Self reported fatigue measured on the Chalder fatigue scale; psychological distress measured on the general health questionnaire, physical functioning and health perception on the SF-36; and a count of physical symptoms.

Results Gulf war veterans experienced a modest reduction in prevalence of fatigue (48.8% at stage 1, 43.4% at stage 2) and psychological distress (40.0% stage 1, 37.1% stage 2) but a slight worsening of physical functioning on the SF-36 (90.5 stage 1, 88.7 stage 2). Compared with the other cohorts Gulf veterans continued to experience poorer health on all outcomes, although physical functioning also declined in Bosnia veterans. Era controls showed both lower incidence of fatigue than Gulf veterans, and both comparison groups showed less persistence of fatigue compared with Gulf veterans.

Conclusions Gulf war veterans remain a group with many symptoms of ill health. The excess of illness at follow up is explained by both higher incidence and greater persistence of symptoms.

Introduction

Consensus exists that service in the 1991 Persian Gulf war resulted in increased symptomatic ill health among those deployed.\(^1\) We know of no studies on the prognosis of symptoms among Gulf war veterans. In 1997 we studied a large random sample of members of the armed forces who served in the 1991 Gulf war,\(^1\) including those who had left the services. We compared the “Gulf cohort” with two military control cohorts. This study assesses the outcomes of these cohorts four years later. Our two main aims were, firstly, to compare the prevalence of various health outcomes over time and between cohorts, and, secondly, to determine rates of incidence and remission for clinically important fatigue and psychological distress after adjusting for potential confounders.

Method

Participants

Our original study consisted of three groups: personnel who served in the Persian Gulf war between 1 September 1990 and 30 June 1991 (the Gulf cohort); personnel who served on UN peacekeeping duties in Bosnia between 1 April 1992 and 6 February 1997 (the Bosnia cohort); and personnel who were serving in the armed forces on 1 January 1991 but who were not deployed to the Gulf (the “Era” cohort).\(^1\) We took a random sample of all Gulf veterans, with oversampling of women. Sampling of the other two cohorts was frequency matched in terms of sex, age, reservist status, officer status, service (Royal Navy, Army, or Royal Air Force), and a measure of fitness.

Of 8196 participants who responded to the first survey 503 refused permission for future contact and 449 failed to complete the relevant section of the questionnaire. We used random stratified sampling to select respondents from stage 1 into the present study. All women were selected. We stratified the sampling on the severity of fatigue at stage 1. The selection process included all male veterans with a fatigue score greater than 8 (511 Gulf, 115 Bosnia, and 120 Era); for Gulf, a 50% sample of veterans with fatigue scores of 4-8 (484 veterans), along with all those in Bosnia (n = 333) and Era (n = 364) who scored in this range; and an approximately one in eight sample of veterans with fatigue scores less than 4 in order to represent asymptomatic individuals (n = 250 in each group).

Mailing method

We used three mailings. To trace non-responders we used the NHS central registry to obtain health authority ciphers and current addresses. We used the online electoral registry “Cameo” to check addresses. Service pension and discharge sources supplied updated addresses. We sent the second and third mailings via commanding officers, asking for their help in disseminating the questionnaires on our behalf. Following an agreement with the War Pensions Agency, the UK Department of Social Security sent two further mailings. In order to comply with data protection regulation, we were not informed which addresses the Department of Social Security had on their records.

Questionnaire and outcomes

The questionnaire included a fatigue scale;\(^1\) the 12 item general health questionnaire (a screening questionnaire for common mental disorders)\(^3\); the SF-36 instrument for physical health and functional capacity\(^4\);\(^5\); and a list of 50 common symptoms. We defined cases of fatigue as having a score on the fatigue scale of greater than 3 and cases of psychological distress as having a score greater than 2 on the general health questionnaire. We defined cases of “stress reaction” from a checklist of symptoms described in previous work.\(^1\)
Results

Statistical analyses

Response bias—We defined four groups—responders, “refusers,” “returns to sender” (where questionnaires were returned to us), and “no information” (where no reply was forthcoming). We compared the frequency of these four outcomes by cohort. We determined whether response bias was present by comparing demographic variables and stage 1 health outcomes across all participants and indicates that Gulf and Era were broadly similar. The Bosnia group were younger, less likely to be married, more likely to have remained in service, and only from the Army.

Table 3 shows the prevalence of categorical outcomes at stages 1 and 2. We report the prevalence of stage 1 outcomes within the sample studied at stage 2, not for the entire cohort—hence the prevalence figures we report for stage 1 are similar, but not identical to, those shown in our previous paper. Table 3 shows that Gulf had higher rates of the disorders under study than the other two cohorts, and this difference is maintained between stages 1 and 2. For Gulf we found a modest reduction in the prevalence of fatigue, post-traumatic stress reaction, general health questionnaire cases, and self reported “Gulf war syndrome.” For Bosnia and Era we found no changes other than an increase in the prevalence of post-traumatic stress reaction in the Era group, which was not significant (P > 0.05).

Table 4 shows the scores on continuous measures at each time point for the three cohorts. The Gulf cohort was less healthy than the other two cohorts at both stages. A decline in physical functioning affected each of the three cohorts (non-significant for Era). Health perception declined for both Bosnia and Era but not younger participants, and those who were unmarried. Non-responders rated their health as poorer at stage 1 for physical disability and general health perception but were less likely to have been cases on the general health questionnaire.

Table 2 shows the sociodemographic characteristics of the participants and indicates that Gulf and Era were broadly similar. The Bosnia group were younger, less likely to be married, more likely to have remained in service, and only from the Army.
For SF-36 scores, negative differences in mean indicate a worsening in health. For other scales, negative scores indicate an improvement in health.

*SF-36 scales range from 0-100, with higher scores indicating better health.

All scores are weighted for sampling.

Table 3  Prevalence of categorical outcomes in the three cohorts. Values are percentages (95% confidence intervals) unless otherwise indicated

<table>
<thead>
<tr>
<th></th>
<th>Gulf</th>
<th>Bosnia</th>
<th>Era</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatigue</td>
<td>Stage 1</td>
<td>Stage 2</td>
<td>Difference</td>
</tr>
<tr>
<td></td>
<td>48.8</td>
<td>42.4</td>
<td>0.45 (0.40-0.50)</td>
</tr>
<tr>
<td>Post-traumatic stress reaction cases</td>
<td>12.4</td>
<td>10.8</td>
<td>0.57 (0.49-0.63)</td>
</tr>
<tr>
<td>General health questionnaire cases</td>
<td>40.0</td>
<td>37.1</td>
<td>0.79 (0.63-0.98)</td>
</tr>
</tbody>
</table>

All prevalence estimates are weighted for sampling.

*Values of <1 indicate declining prevalence. Ratios are weighted for sampling.

Table 4  Scores (95% confidence intervals) for continuous measures by cohort and stage

<table>
<thead>
<tr>
<th></th>
<th>Gulf</th>
<th>Bosnia</th>
<th>Era</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF-36* physical function</td>
<td>Stage 1</td>
<td>Stage 2</td>
<td>Difference</td>
</tr>
<tr>
<td>SF-36* health perception</td>
<td>80.3 (79.0-81.5)</td>
<td>78.7 (77.1-80.2)</td>
<td>-1.6 (1.3-1.9)</td>
</tr>
<tr>
<td>SF-36* health perception</td>
<td>(83.3 to 91.3)</td>
<td>(76.6 to 88.9)</td>
<td>-1.9 (1.6-2.2)</td>
</tr>
<tr>
<td>SF-36* health perception</td>
<td>65.8 (64.4 to 67.7)</td>
<td>65.9 (64.3 to 67.6)</td>
<td>0.0 (1.1-0.3)</td>
</tr>
<tr>
<td>General health questionnaire cases</td>
<td>(14.1 to 14.9)</td>
<td>(13.8 to 14.5)</td>
<td>1.8 (1.1-0.5)</td>
</tr>
<tr>
<td>Fatigue</td>
<td>(17.8 to 17.1)</td>
<td>(16.5 to 17.2)</td>
<td>-2.3 (1.5-3.1)</td>
</tr>
<tr>
<td>Total symptoms</td>
<td>(17.0 to 17.0)</td>
<td>(16.5 to 11.3)</td>
<td>0.0 (0.3-0.7)</td>
</tr>
</tbody>
</table>

All scores are weighted for sampling.

*SF-36 scales range from 0-100, with higher scores indicating better health.

For SF-36 scores, negative differences in mean indicate a worsening in health. For other scales, negative scores indicate an improvement in health.

Table 5  Incidence and persistence of outcomes. Values are presented with 95% confidence intervals

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Risk</th>
<th>Crude odds ratio</th>
<th>Corrected odds ratio*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gulf</td>
<td>20.2 (16.4 to 24.0)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Bosnia</td>
<td>21.2 (16.7 to 25.8)</td>
<td>1.0 (0.7 to 1.5)</td>
<td>0.9 (0.6 to 1.4)</td>
</tr>
<tr>
<td>Era</td>
<td>15.4 (11.4 to 19.4)</td>
<td>0.7 (0.5 to 1.1)</td>
<td>0.7 (0.5 to 1.1)</td>
</tr>
<tr>
<td>Fatigue cases</td>
<td>18.8 (14.4 to 23.1)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Bosnia</td>
<td>19.8 (15.1 to 24.4)</td>
<td>1.1 (0.7 to 1.6)</td>
<td>0.9 (0.6 to 1.5)</td>
</tr>
<tr>
<td>Era</td>
<td>17.2 (13.5 to 21.0)</td>
<td>0.6 (0.3 to 0.9)</td>
<td>0.5 (0.3 to 0.9)</td>
</tr>
<tr>
<td>Post-traumatic stress reaction cases</td>
<td>5.0 (3.8 to 6.4)</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Bosnia</td>
<td>4.0 (2.5 to 5.5)</td>
<td>0.8 (0.5 to 1.3)</td>
<td>0.8 (0.4 to 1.5)</td>
</tr>
<tr>
<td>Era</td>
<td>4.6 (3.0 to 6.2)</td>
<td>0.9 (0.6 to 1.5)</td>
<td>0.9 (0.5 to 1.5)</td>
</tr>
</tbody>
</table>

*Controlled for demographic variables (age, sex, rank, marital status).

Discussion

Main findings

Gulf war veterans continue to experience symptoms that are considerably worse than would be expected in an equivalent cohort of military personnel. However, Gulf war veterans are not deteriorating and do not have a higher incidence of new illnesses.

Our study had two main aims. Firstly, we wanted to describe the outcome of Gulf war veterans three to four years after we first surveyed them. Our results show a disappointing stability in the prevalence of the main disorders we studied. Although the prev
lence of the symptom based disorders lessened for Gulf veterans, physical functioning and health perception measured on the SF-36 barely changed. The reduced physical functioning may have been due to increasing age. The two comparison groups had some worsening in health on the SF-36 scales and more physical symptoms. This implies that some worsening of these health outcomes is expected over time, presumably due to advancing age. At each wave the prevalence between Gulf and the two comparison groups differs, and although the gap narrowed slightly, the Gulf veterans continued to experience poorer health on all measures.

Our second aim was to examine whether the raised prevalence in Gulf veterans was explained by a greater incidence of disorders or more persistence. No clear pattern emerged. We found some evidence that the incidence of fatigue and caseness on the general health questionnaire was higher in the two deployed groups (Bosnia and Gulf) than in Era but that the difference for the general health questionnaire was explained by confounding. For fatigue we found evidence that the Gulf group continued to have greater incidence than the Era group, and this was not explained by confounding. For persistence we found a strong trend for Gulf veterans with fatigue to be more likely to remain fatigued at follow up.

**Limitations of the study**

We achieved a follow up rate of just over 70%, which leaves room for bias. Because follow up rates were worse in participants with poorer health at stage 1 we have probably slightly underestimated the prevalence of the disorders under study. The effect of this bias seems to have been similar across cohorts, which makes it unlikely that the non-response bias would have led to the Gulf group having still higher than expected prevalence figures at stage 1. We believe that the missing values are unlikely to have changed the main findings of this study. We measured health on a range of self report items, which are open to reporting biases.

**Implications**

The nature of Gulf war illness remains ambiguous. If the illness represented the prodrome of a known disease (such as a neurological disorder), even, this has yet to declare itself. We think that the non-specific increase in symptoms reported by our and other studies is likely to remain poorly understood in terms of conventional biomedical diseases. This study did not have the statistical power to assess mortality in Gulf war veterans, and this was not our aim. Other studies in the United Kingdom and the United States, however, have failed to find higher than expected death rates. Finally, as time passes it becomes increasingly difficult to find causes of illnesses in veterans of the 1990-1 Gulf war. We suspect that different psychosocial, military and environmental risk factors may determine onset and recovery, and this is the topic of future research.

We thank Nick Blatchley and Simon Satchell from the Gulf Veterans Illness Unit of the Ministry of Defence for assistance in tracking participants, Michael Dewey for statistical advice, and the participants for their patience in once again completing lengthy questionnaires.

**Contributors:** MH, ASD, and SW designed the study. CU and LH were responsible for data collection under the supervision of SW, ASD, and MH. MH and VN performed the statistical analyses. MH wrote the paper and is the guarantor. All authors provided comments.

**Funding:** US Department of Defense, UK Medical Research Council.

**Competing interests:** None declared.

**Ethical approval:** The study received approval from the relevant research ethics committees.

---

**What is already known on this topic**

Veterans of the 1990-1 Gulf war experience poorer health on most subjective outcomes than non-deployed military personnel.

No satisfactory follow up studies have assessed outcome of veterans of the Gulf war over more than one wave of data collection, so it is unclear whether veterans are getting worse, staying the same, or getting better.

**What this study adds**

Gulf war veterans still have considerably poorer subjective health than appropriate military controls.

The health of Gulf war veterans has improved, but this improvement is relatively minor.

For comparison groups there has been a worsening of health on some outcomes, which is probably due to ageing.

The health gap between Gulf war veterans and comparison groups has therefore narrowed slightly.