Can epidemiology clear the fog of war? Lessons from the 1990–91 Gulf War

Matthew Hotopf* and Simon Wessely

Accepted 19 April 2005

Despite over US $200 million having been spent researching illnesses following the 1990–91 Persian Gulf War, the nature and cause of such illnesses remains controversial. In this narrative review, we discuss some of the methodological issues that have affected epidemiological studies on this topic. These include low-response rates, ascertainment bias, recall bias, problems identifying suitable control groups, and problems defining the outcomes to study. From this we argue that difficulties have arisen partly owing to the significant delay between the point at which illnesses were first identified by veterans and the reporting of epidemiological studies and that health surveillance should be routine following future deployments.

Keywords Gulf war illness, military health

The epidemic of poorly explained multi-symptom illnesses that occurred after the 1990–91 Persian Gulf War, and the alleged birth defects among offspring of veterans continue to pose complex challenges to medical researchers. The US government has spent in excess of US $200 million on a programme of epidemiological and clinical studies. The UK government has supported clinical and epidemiological studies at a cost of £8.5 million.

The main findings of previous epidemiological research are: first that all-cause mortality is no higher in Gulf veterans compared with non-deployed military personnel, however, there is a modest effect of Gulf War service on suicide and accidents. Second, evidence on birth abnormalities among offspring of Gulf veterans is mixed. There was some evidence of increased cardiovascular and urinary malformation in one registry-based study but not in another, a study relying on veterans’ recall found higher rates of birth defects and miscarriages. The best study to date obtained data from the entire Gulf cohort and era controls, albeit with a modest response rate, and found no evidence for a link between fathers’ Gulf War service and the risk of stillbirth, most structural abnormalities, chromosomal malformations, or syndromes. Malformations of the urinary system and non-specific musculoskeletal malformations were elevated, but the risk was reduced when the analysis was restricted to medically verified outcomes only. There was also an increased risk of early (but not late) miscarriage or stillbirths in pregnancies fathered by Gulf veterans. Third, Gulf veterans have greatly increased symptoms and a much higher prevalence of self-reported illnesses than comparable military controls and these tend to persist. Fourth, most but not all controlled studies find that such symptoms do not constitute a unique syndrome.

This paper will attempt to review methodological problems that have affected research on the health of Gulf War veterans. Our aim is not to provide a review of research findings, but instead to point out common problems in the research and how these might be addressed in future work on military health. Some problems we point to are general, and affect all epidemiological studies; others more specifically affect studies of the military. Our aim is to describe how researchers and commissioners of research might act differently after future wars.

History

Reports of increased rates of illnesses apparently associated with the 1990–91 Gulf War were first noted in the year following the end of the campaign, with a claim of unusual illnesses and increased birth defects occurring in Gulf veterans in Mississippi. Later investigations were unable to substantiate these claims but the touch paper had been lit. By 1993 numerous claims were being made in the US, and the same year newspaper accounts began to appear in the UK.

Neither government were prepared for these reports, nor was there any coherent sense of how to approach the growing crisis. Both countries set up an open access referral system for veterans, which attracted large numbers of veterans. However, no discernable pattern could be found in what was being seen, and while it was possible that the individual detailed health assessments reassured individual veterans, the fact that such registries could not be used as a basis for research across the entire veteran or ex-service populations meant that it was impossible for either governments or scientists to extrapolate to the wider community.
It was not until 1996 that research into the health effects of the Gulf War were commissioned in the UK. This delay had many adverse consequences. First, a major opportunity to elucidate the chain of causation was lost. For example, it would have been crucial to know if symptoms developed in both the US and UK at approximately the same time (suggesting a common origin in the circumstances of the war) or at different times (suggesting more cultural/media transmission). This cannot be addressed. As we shall see, the high turnover of the Armed Forces (~5% leave every year), meant that after 6 years a sizeable minority had left the forces, posing major difficulties in tracing and contacting. Linking putative hazards in the Gulf, ascertained many years later was problematic from the start. And finally the field was left open for many years for media speculation, fuelled by the Internet, conspiracy theories, and the occasional maverick scientist. As an example, for most of the 1990s the death of many Gulf War veterans was linked, often in the local media, to their Gulf War service, in the absence of either any other explanation or denominator-based data. The UK only presented mortality data in 2000, showing that there was no increase in mortality in Gulf veterans,26 and cancer data in 2003, showing the same.27 Up to that point, speculation was rife. It would be hardly surprising if anxiety levels rose substantially in the veteran’s community, as more and more reports of deaths accumulated, but without any epidemiological context.

Response rates
Gaining a high participation rate is a key priority for any cross-sectional or cohort study. Studies of military health are at a disadvantage, as the group who in most surveys are least likely to return questionnaires are young men, who make up most of the military population. Those serving in the military are increasingly mobile as the number of deployments has risen, while the size of the UK armed forces has declined, so even those still serving in the military may be difficult to trace. Once they have left the military tracing becomes even more difficult, and there may be suspicion of research that is perceived to be coming from government. Added to this, changes in the implementation of UK data protection law has placed major obstacles for epidemiologists attempting to trace individuals who have not yet consented to participate in research.28

The response rates of Gulf War related research has, with some notable exceptions,17,29 been mediocre. The median rate in systematic reviews is 65%.30,31 What is the likely impact of low rates? The most serious difficulty because of low rates is that individuals who had served in the Gulf had considerably poorer health than our comparison groups.32,33%

Some studies have attempted to test this further by following a proportion of non-responders intensively, in a way which would not be affordable for the entire sample. This approach was used in the King’s Gulf War study.9 Having performed three mailings and achieved a response rate of 65%, we attempted to trace 200 non-responders by using more intensive methods, such as finding addresses from the Driving and Vehicle Licensing Agency, contacting family doctors, and using shortened telephone interviews. We successfully gained information on 139 of these individuals, and found that there were important demographic differences between the groups (they were somewhat older, less educated, and less likely to have remained in the Armed Forces). We also found differences in health outcomes, with the intensive follow-up group having poorer perceived health and physical functioning. However, these differences were slight, and our conclusion was that non-response bias was unlikely to radically alter our main finding, namely that individuals who had served in the Gulf had considerably poorer health than our comparison groups.

Another approach is to compare early responders with late responders. Late responders in a survey with a high participation rate could, perhaps, be considered similar to non-responders in surveys with lower participation rates. A cross-sectional survey of civilians, which aimed to determine the prevalence of dependence owing to disablement,32 found an interaction between age and dependence: among older participants (>65 years), late responders were identical to early responders in terms of dependence. However, for younger adults, the disabled group replied earlier, and prevalence apparently fell as the survey continued. This is an important finding, since it indicates that for adults of working age, illness may lead to higher response rates, presumably reflecting the occupational impact of illness in younger groups. We have demonstrated a similar effect in Gulf veterans,33 in that early responders do indeed have more symptoms than later responders. Taking account of this reduces, but does not eliminate, the health effect of deployment to the Gulf.

These findings should to some extent be reassuring—low participation rates are associated with higher symptom reporting, and as studies improve their participation rates, the ‘Gulf effect’ would therefore be slightly reduced, but this bias cannot account for the large effect sizes which have been reported. A few caveats are needed here, though. First, any attempt at intensive follow-up, or comparison of early and late responders works on the assumption that the ‘never responded’ are similar to late responders—there is, however, a risk that those who never respond, even to intensive follow-up, are not on the same continuum, possibly consisting of a nucleus of individuals with particularly poor outcomes. Second, whilst low participation rates may not have mattered for Gulf War surveys,
this conclusion cannot be extrapolated to other settings. Low participation rates may not have mattered in studies of Gulf War veterans, because there was a consistent and sizeable effect, which was unlikely to be accounted for by bias. It would be an over-generalization to conclude that the same will apply for studies of recent or future deployments.

**Other sources of selection bias**

Ideally, all those who are deployed to a theatre of war should be followed, but many studies on the Gulf War limited research to personnel who were still serving in the military. Such studies, many of which were conducted several years following deployment, are likely to exclude many of the people in whom we are most interested. The symptomatic may be the most likely to leave the military, and may also be most disaffected and alienated. Failure to include such personnel in surveys may serve to over-estimate the health of participants, leading to selection bias.

**Ascertainment bias**

Another problem encountered in studies of Gulf War related health comes from ascertainment bias. First, as we have discussed response rates tend to be higher in Gulf veterans than non-deployed comparison groups. Second, efforts to find Gulf War veterans are often conducted with more vigour than in the controls. In particular, the existence of Gulf War registries in the US, but not equivalent registries for controls, may cause problems. As an example, to date the only well-defined condition for which the US government has accepted a link with Gulf War service is motor neuron disease (MND), known as amyotrophic lateral sclerosis (ALS) in the US. This was on the basis of a series of distressing reports of MND in Gulf veterans, which led to the then commissioning of epidemiological research. When eventually published, some 2 years after service linkage was accepted by the government, the rates of MND appeared higher in the Gulf veterans than controls. Numbers of course were small, since MND is fortunately a rare disease in all populations. The rate ratio for MND comparing all deployed and non-deployed was 1.92 (1.29–2.84) with 40 cases in the deployed group, compared with 67 in the non-deployed. The problem is that it seemed to have been easier to find cases in the Gulf sample than the controls. Although the authors attempted to adjust for this, and found an incidence of MND similar to that of the general population in the non-deployed group, the suspicion of ascertainment bias remains because there has been no change in overall mortality, yet MND is a rapidly fatal disease. Mortality statistics are unlikely to be subject to ascertainment bias. Unless there are uniform surveillance systems in place for all the military, not just those involved in deployments that subsequently become controversial, suspicions that modest increases in the risk of rare diseases are owing to ascertainment bias will always be present.

**Ascertainment of exposures**

One major theoretical advantage of classic cohort studies is that they follow individuals from the point of an exposure to the development of an outcome and beyond. This temporal sequence aids the interpretation of the relationship between an exposure and an outcome, as it is possible to ascertain whether the exposure preceded the outcome. While many studies of Gulf War veterans have been cohort studies, in the sense that they defined participants according to their exposure (i.e. deployment), they are historical or retrospective cohort studies, in that participants were recruited long after the deployment happened. In effect these studies are more like cross-sectional studies where parallel populations (deployed/not deployed) are studied at a single point in time. The exact definition of the study design matters less than the effect the design has on the interpretation of findings.

No one can dispute that military personnel deployed to the Gulf were exposed to a wide range of potentially harmful exposures. The exposures that have received most attention relate to biological or chemical weapons, and the precautions taken against these. While there is no evidence that Iraq used chemical weapons, multiple chemical weapons alarms were sounded, and pyridostigmine bromide (the prophylaxis used against nerve agents) was consumed. Similarly, vaccination programmes included routine vaccines (e.g. typhoid and tetanus) as well as vaccines to biological agents such as anthrax, plague, and botulism. Pesticide and insect repellent use was widespread, and included organophosphates in low doses and DEET (the insect repellent N,N-diethyl-meta-toluamide). Depleted uranium (used in anti-tank shells) has received considerable interest in the media, but exposure was rarely reported in our survey. Some of the most common exposures have received less interest, but include smoke from oil well fires and exposure to diesel or petrochemical fumes or fuels, as well as potentially traumatic and distressing experiences such as seeing the dead, wounded and maimed, or handling prisoners of war.

The degree of exposure to such hazards clearly varied by time, place, and person. Uptake of vaccine programmes in UK armed forces varied considerably between units. We can be reasonably confident of the point in time when some of the precautions against biological and chemical weapons were likely to be used. Front line personnel involved in combat would clearly have had different exposures to some hazards (e.g. depleted uranium, where exposure is only likely in those who are in close proximity to spent munitions or clear-up operations) compared with support personnel, who may have been many miles away from areas where war fighting was taking place. Unfortunately, while careful efforts have been made after the event to characterize the patterns of exposures to hazards that might have been expected according to units, such exercises are imprecise. Ascertainment of exposures has relied heavily on the recall by veterans.

Many studies of Gulf veterans’ health have attempted to measure deployment-related exposures. Hence our study demonstrated that reporting multiple vaccines was associated with self-reported illness, a finding replicated by two other groups. Our study also showed that practically every exposure we asked about was associated with every outcome we measured, a pattern demonstrated by others. Exposure to military hazards is not random. Personnel exposed to one hazard may be more likely to be exposed to another (although in theory this should not apply to medical counter-measures given to all personnel in theatre). Nevertheless, hearing chemical alarms, having combat exposure, and being near depleted uranium
munitions will be correlated, creating problems of multicolinearity. This clustering of exposure data should also remind us that—despite the growing complexity of the modern military—deployment typically occurs in clusters, whereby groups of individuals are deployed to carry out similar duties, in similar geographical areas and are, therefore, likely to encounter similar hazards. Furthermore, such groups are social units, and certain psychosocial variables (e.g. morale) are best considered to belong to the group rather than the individual. To disentangle this, study designs and analyses should ideally take account of the grouped nature of data, ideally using multi-level models. To our knowledge no study of deployment to the Gulf has done so.

Our finding that most exposures were associated with most outcomes also raises the question of recall bias. Recall between those who are ill and those who are well may differ for many reasons, other than a true causal association between exposures and outcomes. Those who are well may be more inclined to forget, or at least down-play, exposures. Those who are ill may ruminate over their experiences. Furthermore, frequently repeated narratives heard on the media may reinforce inaccurate views about exposure. After the Vietnam War, many US soldiers gave compelling accounts of being stained by Agent Orange, used to defoliate jungle, but also seen as a serious hazard blamed for poorly explained illnesses. Compelling though such accounts were, they could not be substantiated—Agent Orange is a code name for a chemical which is colourless and does not stain the skin.

Relatively few studies have attempted to quantify the potential for recall bias. McCauley et al. used an ingenious method where three approaches were used. Using a survey of US personnel who had been deployed to the Gulf, the researchers were able to classify them according to whether they were deployed exclusively before, during, or after the war. The patterns of exposures between these groups would have been expected to vary—pyridostigmine bromide was not issued in the months prior to combat, similarly SCUD alarms would only have been heard by individuals during combat. However for some hazards a sizeable minority (11.6% for pyridostigmine bromide, 21.7% for hearing SCUD alarms) of veterans who had left the theatre before the exposure happened, indeed reported them. A remarkable finding was that 28% of those deployed left the theatre before the exposure happened, indeed reported experiencing it.

How can these measurement issues be overcome? First, wherever possible, objective measures of exposure should be used. For example, we were able to identify a subgroup of individuals who had their own vaccine records. This group of individuals were able to provide invaluable information on what they had received, and reassuringly, results linking multiple vaccine exposure to illness were broadly similar in this subgroup to that of the remaining participants. Second, there may be useful clues about the possible impact of individual exposures from the descriptive epidemiology of Gulf War illness. By descriptive epidemiology, we mean the traditional epidemiology that may help eliminate some putative risk factors. Furthermore, we were able to show that changes in recall were non-random. Individuals who reported fewer exposures at time 2 than they did at time 1, tended to report an improvement in their perceived health. Similarly, reporting more exposures at time 2 compared with time 1 was associated with a worsening in health. Thus, at least to some extent, there is evidence of recall bias.

Our finding that most exposures were associated with most outcomes also suggests how these measurement issues may be overcome. McCauley et al. linked dates of response to the survey to media stories publicizing new putative risk factors. The rates of reporting of insect repellent cream did increase in response to news coverage, but the effect was modest and did not apply to all exposures examined. Finally, they assessed test–retest reliability of reports of exposures in the same individuals at two points in time, and found a wide range of kappa values ranging from 0.07 to 0.86 with a median of 0.53. We have replicated this last finding by using a follow-up of our initial survey. Respondents were asked to complete the same questionnaire 2–4 years apart. Kappa values, indicating the degree to which reports of exposures were consistent above chance, ranged from 0.07 to 0.79, with most exposures falling into the range from 0.35 to 0.55, indicating only modest test–retest reliability.

These findings indicate that there is scope for considerable inaccuracy of recall, and on the basis that a variable unreliably measured cannot be valid, these findings must cast considerable doubt on the ability of surveys to describe accurate associations between exposures and outcomes. Furthermore, we were able to show that changes in recall were non-random. Individuals who reported fewer exposures at time 2 than they did at time 1, tended to report an improvement in their perceived health. Similarly, reporting more exposures at time 2 compared with time 1 was associated with a worsening in health. Thus, at least to some extent, there is evidence of recall bias.

Our finding that most exposures were associated with most outcomes also suggests how these measurement issues may be overcome. McCauley et al. linked dates of response to the survey to media stories publicizing new putative risk factors. The rates of reporting of insect repellent cream did increase in response to news coverage, but the effect was modest and did not apply to all exposures examined. Finally, they assessed test–retest reliability of reports of exposures in the same individuals at two points in time, and found a wide range of kappa values ranging from 0.07 to 0.86 with a median of 0.53. We have replicated this last finding by using a follow-up of our initial survey. Respondents were asked to complete the same questionnaire 2–4 years apart. Kappa values, indicating the degree to which reports of exposures were consistent above chance, ranged from 0.07 to 0.79, with most exposures falling into the range from 0.35 to 0.55, indicating only modest test–retest reliability.

These findings indicate that there is scope for considerable inaccuracy of recall, and on the basis that a variable unreliably measured cannot be valid, these findings must cast considerable doubt on the ability of surveys to describe accurate associations between exposures and outcomes. Furthermore, we were able to show that changes in recall were non-random. Individuals who reported fewer exposures at time 2 than they did at time 1, tended to report an improvement in their perceived health. Similarly, reporting more exposures at time 2 compared with time 1 was associated with a worsening in health. Thus, at least to some extent, there is evidence of recall bias.

Our finding that most exposures were associated with most outcomes also suggests how these measurement issues may be overcome. McCauley et al. linked dates of response to the survey to media stories publicizing new putative risk factors. The rates of reporting of insect repellent cream did increase in response to news coverage, but the effect was modest and did not apply to all exposures examined. Finally, they assessed test–retest reliability of reports of exposures in the same individuals at two points in time, and found a wide range of kappa values ranging from 0.07 to 0.86 with a median of 0.53. We have replicated this last finding by using a follow-up of our initial survey. Respondents were asked to complete the same questionnaire 2–4 years apart. Kappa values, indicating the degree to which reports of exposures were consistent above chance, ranged from 0.07 to 0.79, with most exposures falling into the range from 0.35 to 0.55, indicating only modest test–retest reliability.
Comparison groups

Finding suitable comparison groups for occupational cohort studies is a common challenge for epidemiologists, because of threats to internal validity such as the healthy worker effect.\(^57\) When the occupation under study is service in the military the issue becomes more complicated for at least two reasons. First, the military is a closed organization. In UK recruitment has been strongest from areas of high economic deprivation. The military has its own unique class structure that is less flexible than in civilian life and strongly confounded by education and family background.\(^58\) Much of the focus on military training is exercise and fitness; hence, members of many branches of the military may be especially fit. However, perhaps reflecting the appeal of the military for many recruits, risk taking behaviours are higher than in comparable socioeconomic groups. This is reflected by high rates of smoking and alcohol consumption.\(^59\) For all these reasons it would be inappropriate to compose a non-military comparison group for studies on military health.

The obvious solution used in many of the main studies on Gulf War illness, was to compare military groups who were deployed to the Gulf with others who were not. However, even with a military comparison group, a further complication comes into effect namely the health warrior effect.\(^60\) This is essentially the same as the healthy worker effect, but reflects the problem that those who are in the military who are ill are less likely to be deployed. It is perhaps a greater potential problem for US forces, having a higher proportion of reservists, and (at least before September 2001) were relatively infrequently deployed. In contrast UK forces, where the majority of those deployed were in the regular forces, have experienced very heavy deployment duties over the last 15 years.

How much of a problem is the healthy warrior effect? For the many studies that show a powerful relationship between deployment and symptoms, the effect is more a theoretical nicety than a serious concern. It is a greater problem for studies of mortality, cancer registrations, or hospitalization following deployment, which have tended to be negative. In a themed edition of the *American Journal of Epidemiology*, the criticism was made that these ‘negative’ findings were dubious because personnel who were not deployed were more likely to be suffering from HIV, cancer, or other chronic diseases.\(^61\) While this theoretical concern exists, researchers have pointed out that the military are in general a young and fit workforce, where rates of such chronic diseases are low.\(^62\) When put to the test directly, there was no evidence that military personnel deployed to theatres other than the Gulf had lower mortality than those not selected for deployment.\(^63\) Gray *et al.*\(^64\) assessed whether healthy warrior effect could have affected their findings on hospitalizations\(^65\) following deployment to the Gulf. Their main finding was that rates of hospitalization fell for the Gulf group just before and during deployment. Their interpretation was that this was most likely owing to operations for minor complaints (e.g. excision of lipomas) being postponed in the deployed group.

One way to overcome the healthy warrior effect is to compare two or more deployed groups. We compared Gulf veterans with both non-deployed military personnel and personnel who had been deployed to Bosnia on UN peacekeeping duties in 1992–96.\(^9\) This design was unique among studies on the general health of Gulf War veterans. It was reassuring to find that those deployed to Bosnia behaved very much like the non-deployed comparison group. However, using a deployed comparison group also causes potential problems. Each deployment has its own unique stresses. Peace-keeping deployments such as Bosnia may give the impression of being less stressful than war-fighting, and to be associated with fewer hazards, but it is safer to say that the stresses and hazards of each deployment are different. Personnel in Bosnia were more likely to come under small arms fire than the Gulf group, but unsurprisingly were less likely to hear chemical weapons alarms and to believe themselves to have experienced a chemical alert.\(^66\)

Our current experience of defining a comparison group for individuals who have been deployed to Operation Telic (the 2003 Iraq War) illustrates that for future studies of post-deployment health, defining appropriate comparison groups may be more difficult. The continued presence of British Armed Forces in Iraq means that a sizeable proportion of the comparison group have now been deployed to Iraq. Clearly such individuals cannot be re-assigned to the deployed group, since their experiences will be very different from the war-fighting at the start of the Iraq war. However, leaving them in the comparison group may attenuate any health effect of deployment, as these individuals will arguably have experienced a more stressful period of the engagement. However, taking such individuals out of the comparison group may lead to a selection bias, especially if the total proportion of the comparison group who are deployed to Iraq on later operations becomes sizeable.

Outcomes

A major problem for Gulf War health studies was choosing the outcome to be studied. From the beginning, Gulf War veterans reported a complex constellation of physical symptoms, but studies on mortality and hospitalizations did not indicate an increase in deaths or occurrences of well-defined biomedical diagnoses.\(^3,27\) Clinical studies that have been performed since have found no more biomedical diagnoses than would have been expected by chance.\(^67–69\) What we are left with is an epidemic of symptoms, not diagnoses.

The first and most obvious conclusion is that while there may not have been an increase in diseases previously recognized to biomedicine, deployment to the Gulf War had caused a new disorder. The presence of a Gulf War syndrome is perhaps the most highly contentious issue in the current controversy. While apparently of great importance in political terms, the issue is resolvable from a scientific perspective. Figure 1 shows the distribution of symptoms in our study comparing Gulf, Bosnia, and era controls.\(^9\) To the left are the most common symptoms in each group (e.g. headache and fatigue) and to the right are the least common ones. A similar general pattern of increased symptom reporting in Gulf veterans has been described in other studies.\(^17,46\) Three simple conclusions can be drawn from this graph. First, the group who went to the Gulf had approximately double the rate of each symptom compared with the two comparison groups. Second, the pattern of symptoms was similar across the three groups. If a new syndrome was present, one would expect a specific increase in some symptoms, with
logic plausibility, are dependent upon the initial assessment and methods are limited in that they do not address issues of bio-
sophistry. Further, as Drue Barrett of symptoms seen in Figure 1 provides the same answer with less
formally test hypotheses, a simple description of the distribution
multivariable techniques such as confirmatory factor analysis can
structure unique to the Gulf. We suggest that while complex
statistical study to address this using factor analysis was
increase in symptom burden. Likewise, comparisons between
establishing a single case definition if the outcome is a general
group. Third, there are going to be considerable difficulties in
the remainder being experienced at the same rates across each
group. Third, there are going to be considerable difficulties in
establishing a single case definition if the outcome is a general
increase in symptom burden. Likewise, comparisons between
studies are also going to be problematic.

The initial conclusion to emerge from the first formal
statistical study to address this using factor analysis was
different. Haley et al.70 studied 249 Gulf War veterans selected
from a single unit. The study had a response rate of 41%. They
reported six factors labelled ‘impaired cognition’ (characterized
by self-reported distractibility, short-term memory problems,
and depression), ‘confusion-ataxia’ (characterized by self-
reported confusion, difficulties with balance, or diagnosis of
depression or post-traumatic stress disorder) and so on. Here
then was the new Gulf War syndrome, or at least syndromes.
However, the report ignored the fact that in any population one
would expect constellations of symptoms to group together,
and these groups would show up as separate factors in a
factor analysis. This technique has been used in psychometrics
to define variability of personality traits,71 or emotional
symptoms.72 The existence of groups of symptoms identified in
factor analysis says nothing about new syndromes—only that
symptoms tend to group together.

The problem can only be resolved by using comparison groups
of non-deployed military personnel, and comparing factor
structures across groups. If a new, unique, Gulf War syndrome
was present, one would expect those deployed to the Gulf to
show a different factor structure to those not deployed.
Confirmatory factor analysis does just this, and all3–75 but one76
study that has used this technique have failed to find a factor
structure unique to the Gulf. We suggest that while complex
multivariable techniques such as confirmatory factor analysis can
formally test hypotheses, a simple description of the distribution
of symptoms seen in Figure 1 provides the same answer with less
 sophistry. Further, as Drue Barrett et al. point out ‘data-analytic
methods are limited in that they do not address issues of bio-
logic plausibility, are dependent upon the initial assessment and
inclusion in the analysis of the appropriate symptoms, and
involve subjective interpretation of the factors’30.

What if we accept that a new syndrome formally defined by
statistical methods does not exist? Should researchers still aim
to describe a new case definition based upon the most of the
veterans’ complaints, or should existing diagnostic frameworks
be used? To answer this, it is necessary to decide whether
existing diagnostic frameworks provide diagnostic labels for a
sufficient proportion of sick veterans. There is no shortage of
diagnostic labels for symptom-based conditions. Most of these
have been derived by expert panels seeking to ensure
consistency of diagnosis (e.g. the CDC criteria for chronic
fatigue syndrome).77 We have previously argued that different
branches of medicine have responded to the need to describe a
similar group of patients presenting in each specialty’s clinic by
creating symptom-based conditions.78 For example, rheuma-
tologists require a diagnosis for patients with musculoskeletal
symptoms without clear biomedical explanation, and
fibromyalgia has fulfilled that role. Irritable bowel syndrome
and ‘syndrome X’ (or non-cardiac chest pain) fulfil the same
role for gastroenterology and cardiology, respectively. While
there is considerable controversy over the aetiology of these
conditions, they all share the feature of prominent and disabling
physical symptoms for which no adequate current biomedical
explanation exists. Despite these similarities and considerable
overlap between syndromes, the number of diagnostic labels
has risen with increasing sub-specialization in medicine, and
the lack of a unifying approach to the problem has tended to
limit research in this clinically important area.79 Psychiatric
classification does little to help here. The ‘somatoform disorders’
refer to conditions in which physical symptoms predominate
but for which no biomedical explanation can be found.80 The
classification of such disorders is based more by presumed (but
unproven) psychopathological processes than readily applied
symptom clusters, and the approach has been criticized for its
arbitrariness. The most helpful approach might also be the most
simple, in which patients are classified as being a ‘case’ on the
basis of having a certain number of medically unexplained
symptoms.

Inasmuch as they have multiple medically unexplained
symptoms, Gulf War veterans do not necessarily present with a
new syndrome but with high rates of syndromes that have
already been described. One clinical study evaluating help-
seeking veterans found 16% reported symptoms consistent
with chronic fatigue syndrome, and 13% symptoms consistent
with multiple chemical sensitivity.81 In another study, 36% of
symptomatic veterans who reported fatigue and chemical
sensitivity met the criteria for chronic fatigue syndrome, and
13% symptoms consistent with multiple chemical sensitivity.82
Elevated rates of these symptom-based conditions have also
been reported in several cross-sectional studies of Gulf War
veterans.11,14,83 Hence diagnostic groupings that describe many
of the symptoms reported by Gulf veterans exist, but owing to
the Cinderella status of symptom-based research, has not been
widely adopted by researchers into Gulf War illnesses.

An alternative approach is to generate a new diagnostic
grouping in order to describe Gulf War illnesses. This approach
was adopted by Fukuda,11 who, having performed a factor
analysis to describe the main symptom groups, derived a case
definition in which one or more symptoms had to be present

![Figure 1 Frequency of common symptom in Gulf War veterans compared with two military control groups. Reproduced from Unwin et al.79](https://example.com/figure1.png)
in two of three categories: fatigue, mood and cognition, or musculoskeletal symptoms. While this approach does describe many symptomatic individuals (Fukuda's study estimated prevalence at 47% for veterans, and 15% for the non-deployed; our study using a less well-defined definition based on the same symptoms showed a prevalence of 62 and 32% for deployed and comparison subjects, respectively) it also indicates a lack of specificity—you do not have to go to the Gulf War to develop a ‘Gulf War illness’, as might have been expected from the preceding discussion on data-derived syndromes.

New diagnostic labels have obvious benefits in allowing researchers and clinicians to apply consistent criteria and enhancing a common language. However, they also have disadvantages. Labels based on no more than meetings of experts have no criterion validity, and are unlikely to ‘carve nature at the joints’. An example is the CDC diagnostic criteria for chronic fatigue syndrome, which as well as requiring the nature at the joints’. An example is the CDC diagnostic criteria.

References


Ministry of Defence. Further memorandum concerning the provision of advice to MOD Ministers between 1994 and 1996 on the subject of Gulf War Illnesses; Pyridostigmine Bromide. CT-164


Ministry of Defence. A Review of the Scientific Literature as It Pertains to Gulf War Illnesses; Pyridostigmine Bromide. CT-164


Ministry of Defence. A Review of the Scientific Literature as It Pertains to Gulf War Illnesses; Pyridostigmine Bromide. CT-164

Ministry of Defence. A Review of the Scientific Literature as It Pertains to Gulf War Illnesses; Pyridostigmine Bromide. CT-164

Ministry of Defence. A Review of the Scientific Literature as It Pertains to Gulf War Illnesses; Pyridostigmine Bromide. CT-164

Ministry of Defence. A Review of the Scientific Literature as It Pertains to Gulf War Illnesses; Pyridostigmine Bromide. CT-164

Ministry of Defence. A Review of the Scientific Literature as It Pertains to Gulf War Illnesses; Pyridostigmine Bromide. CT-164

Ministry of Defence. A Review of the Scientific Literature as It Pertains to Gulf War Illnesses; Pyridostigmine Bromide. CT-164


CIA and DoD. Modelling the chemical warfare agent release at the Khamisihay pit. 1997.


