Screening for physical and psychological illness in the British Armed Forces: III: The value of a questionnaire to assist a Medical Officer to decide who needs help

R J Rona, R Hooper, M Jones, C French and S Wessely

OBJECTIVES: To estimate the positive and negative predictive values (PPVT and NPVT), sensitivity and specificity of a full and abridged screening questionnaire of physical and psychological health, using primary care doctors’ (medical officers [MOs]) assessments as to whether the servicemen needed medical help as a gold standard.

METHODS: From a tri-service random sample of those who completed a questionnaire, all ‘screen-positive’ and an equal random sample of ‘screen-negative’ were selected to attend their medical centre. MOs were aware that the screening was aimed at detecting psychological illness, but were blind as to the ‘screen-positivity’ of any serviceman. The MO completed a questionnaire that asked whether the patient needed medical help and whether s/he was previously aware of this need.

RESULTS: 314 subjects were available for analysis. The PPVT was 47% (95% confidence interval [CI] 36–59%) for the full questionnaire and 48% (95% CI 36–60%) for the abridged questionnaire. Of those ‘screen-positive’ subjects whom the MO rated as needing help, one third had problems already known to the MO, regardless of the length of the questionnaire. The sensitivity and specificity of the full and abridged questionnaires were 43% and 74%, and 36% and 83% respectively. The PPVT did not vary greatly between health dimensions nor did selection of servicemen with very high scores.

CONCLUSIONS: The use of MOs as a gold standard is important because of their central role in initiating the management of any condition uncovered by a screening programme. Using MOs as a gold standard, the validity of the screening questionnaires for physical and psychological health in the military was mediocre.

The USA and Australia have supported the introduction of screening for physical psychological illness in the military despite the lack of evidence of the value of such screening.1-3 The Health Surveillance Steering Group (HSSG), answerable to the UK Surgeon General’s Department, asked us to assess the validity and eventually the effectiveness of a screening programme. The study would be carried out as a pragmatic project that would reproduce the conditions prevailing on the ground. A key element in the success of this kind of screening is that the medical officers (MOs) would act upon a referral triggered by a screening questionnaire.

So far most studies have concentrated on GP recognition of patients with psychiatric illness after pen and paper self-assessment tests, but little is known about the determinants of the help-seeking behaviour in those with a psychiatric illness not presenting to a GP.4,5 Meta-analysis on the value of the help-seeking behaviour in those with a psychiatric illness not presenting to a GP.4,5 Meta-analysis on the value of the help-seeking behaviour in those with a psychiatric illness not presenting to a GP.

Screening questionnaire

The full questionnaire included the PTSD checklist,12 the General Health Questionnaire 12 (GHQ-12),13 15 physical symptoms,10 a self-assessment of health status from the SF-36 [AQ1]14 and three modified questions on alcohol behaviour.15 A description of the abridged questionnaire and referral criteria are given elsewhere.11

METHODS

Sample

A total of 4500 servicemen were chosen by first randomly selecting units in the three Services and then by randomly selecting 45 servicemen in each unit. All servicemen in the same unit received the same type of questionnaire.11 All medical centres with which the selected servicemen were registered participated in the study.
Referral questionnaire
The post-consultation questionnaire completed by the MO was used to assess the validity of the questionnaire based on the MO’s views as to whether the servicemen needed medical help and, if appropriate, whether the MO was already aware of the serviceman’s health problem.

The screening procedure
‘Screen-positive’ and ‘screen-negative’ were identified by two research assistants according to the score on each health dimension. All ‘screen-positives’ and a random sample of ‘screen-negatives’ were invited to attend their medical centre. The ratio of the two groups was one to one.

Medical centres were advised about the study by letter, telephone calls and in the Surgeon General’s Newsletter. We approached the senior MO and practice manager of the 130 participating medical centres. Medical centres received an explanatory letter and a list of referred servicemen, but we did not disclose whether an individual was a ‘screen-positive’ or a ‘screen-negative’. The letter to the doctor stated that the screening questionnaire asked questions on symptoms, life style and psychiatric illness, including PTSD. We asked the doctor to make a general assessment of the subject’s state of health, focusing on any concerns that the subject might raise. It also explained why the doctor would be blind as to whether a serviceman was ‘screen-positive’. A copy of the screening questionnaire was also sent.

Assessment of validity
Based on the MO’s responses we ascertained the validity of both the full and abridged questionnaires. In our study, the numbers of ‘screen-positive’ and ‘screen-negative’ were fixed by design. Thus it was necessary to weight these groups to reflect the proportions found in the population of servicemen before calculating sensitivity and specificity. The positive predictive value of the test (PPVT) should give an indication for the percentage of servicemen that were identified as ‘screen-positive’ who were confirmed as needing medical help by the MO. The MOs did not receive any specific training in relation to the aims evaluated and were blind to the reason for the referral. Our results represent unmodified practice. We assessed validity for at least one dimension above the threshold and for each health dimension. We also assessed the validity of the questionnaire if only ‘screen-positives’ with extreme scores were referred. For such an analysis, only carried out in those who completed the full questionnaire, a ‘screen-positive’ had one of the following: a GHQ-12 score of eight or over; eight or more symptoms; five or more moderate or severe symptoms; two or more severe symptoms; PTSD of 50 or over; or any two of GHQ, symptoms and health perception above the original cut-offs [AQ2]. In this analysis alcohol behaviour was not included.

RESULTS
There were data from 177 and 137 servicemen available for analysis from the full and abridged questionnaires respectively. MOs were of the opinion that 47% of ‘screen-positive’ and 30% of ‘screen-negative’ servicemen on the full questionnaire, and 48% of ‘screen-positive’ and 25% of ‘screen-negative’ servicemen on the abridged questionnaire needed medical help. The PPVT and negative predictive value of the test (NPVT) and the sensitivity and specificity, with 95% CI, were similar for the two questionnaires (Table 1). The screening questionnaire was not a good tool for detecting whether an MO thought a patient needed help.

The percentage of patients who needed help according to the doctor and of whom s/he was previously unaware was higher in ‘screen-positive’ than ‘screen-negative’, but the difference was relatively minor. 11% for the full questionnaire and 13% for the abridged questionnaire.

There was no evidence that the PPVT was better for some health dimensions than others, although confidence intervals were wide (Table 3). As ‘screen-negative’ for a

Table 1: PPVT and NPVT, and sensitivity and specificity of the full and abridged screening questionnaires (95%CI)

<table>
<thead>
<tr>
<th></th>
<th>Full questionnaire</th>
<th>Abridged questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% (95% CI)</td>
<td>% (95% CI)</td>
</tr>
<tr>
<td>PPVT</td>
<td>47% (36-59%)</td>
<td>48% (36-60%)</td>
</tr>
<tr>
<td>NPVT</td>
<td>70% (60-79%)</td>
<td>75% (62-85%)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>43% (31-55%)</td>
<td>38% (22-49%)</td>
</tr>
<tr>
<td>Specificity</td>
<td>74% (67-82%)</td>
<td>83% (78-89%)</td>
</tr>
</tbody>
</table>

NPVT, negative predictive value of the test. PPVT, positive predictive value of the test.

Table 2: Doctors’ awareness of servicemen’s needs for medical help according to servicemen’s status in the screening questionnaire

<table>
<thead>
<tr>
<th></th>
<th>Doctor’s classification</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Needed help and doctor previously aware</td>
<td>Needed help and doctor not previously aware</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Result of screening</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>40 (53%)</td>
<td>12 (16%)</td>
<td>24 (32%)</td>
<td>76 (100%)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>71 (70%)</td>
<td>9 (9%)</td>
<td>21 (21%)</td>
<td>101 (100%)</td>
</tr>
<tr>
<td></td>
<td>Weighted average*</td>
<td>65%</td>
<td>11%</td>
<td>24%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>Result of screening</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>38 (52%)</td>
<td>12 (16%)</td>
<td>23 (32%)</td>
<td>73 (100%)</td>
</tr>
<tr>
<td></td>
<td>Negative</td>
<td>48 (75%)</td>
<td>4 (6%)</td>
<td>12 (19%)</td>
<td>64 (100%)</td>
</tr>
<tr>
<td></td>
<td>Weighted average*</td>
<td>70%</td>
<td>9%</td>
<td>22%</td>
<td>100%</td>
</tr>
</tbody>
</table>

* Weighted to allow for different sampling fractions of screen-positive and screen-negative; i.e. expected result if entire population is referred.

Journal of Medical Screening 2004 Volume 11 Number 2 www.jmedscreen.com
The validity of screening

Table 3

<table>
<thead>
<tr>
<th></th>
<th>PPVT of screening questionnaires by health criterion of referral (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Criterion of Referral</strong></td>
<td><strong>PPVT (95% CI)</strong></td>
</tr>
<tr>
<td><strong>Full questionnaire</strong></td>
<td></td>
</tr>
<tr>
<td>Symptoms</td>
<td>47% (31-64%)</td>
</tr>
<tr>
<td>GHQ</td>
<td>50% (34-66%)</td>
</tr>
<tr>
<td>Alcohol</td>
<td>42% (21-66%)</td>
</tr>
<tr>
<td><strong>Abridged questionnaire</strong></td>
<td></td>
</tr>
<tr>
<td>Symptoms</td>
<td>29% (10-58%)</td>
</tr>
<tr>
<td>GHQ</td>
<td>51% (38-63%)</td>
</tr>
<tr>
<td><strong>Full and abridged questionnaire</strong></td>
<td></td>
</tr>
<tr>
<td>PTSD</td>
<td>67% (24-94%)</td>
</tr>
</tbody>
</table>

*Too few servicemen attended the medical centre related to the criterion of self-assessment of health for useful analysis. PPVT, positive predictive value of the test.

single health dimension would include individuals who are ‘screen-positive’ for other health dimensions it was inappropriate to estimate NPVT for each dimension. It was not possible to work out sensitivity and specificity for individual dimensions because MOs were not asked to give diagnosis on each dimension separately.

If we had selected as ‘screen-positive’ only those with extreme scores, as defined in the methods section, 11.4% of the servicemen would have been identified as ‘severe screen-positive’ and 20.3% as ‘mild screen-positive’ from the total sample of 2890 subjects who completed the screening questionnaire. From our results we estimate that the PPVT would have been 54%, the NPVT 67%, sensitivity 17% and specificity 74%. A change of threshold would not have greatly changed the validity of the full questionnaire.

**DISCUSSION**

This study demonstrates that the validity of a pen and paper questionnaire when the gold standard is an MO is mediocre. There was a high degree of misclassification between the questionnaires and the MO’s assessment, regardless of length of the questionnaire and the threshold of each health dimension.

We cannot answer the question whether another approach to screening, e.g. using psychologists as the gold standard, would have been more valid, but we suspect that the validity would be similar as many of the problems we encountered with the screening process would have been similar. The British Armed Forces has a periodic examination, the PULHHEEMS, that has a psychiatric component, but it would be fair to state that it yields little information and it is very rare for a diagnosis to be made as a result of such an assessment.

The voluntary nature of our study led to a low response rate of uptake of consultations with MOs and we commented on this feature in the accompanying paper.11 We can speculate that if the participation in the study were made compulsory the validity would not have improved. Many servicemen would not be prepared to volunteer symptoms or illness because their medical recognition might jeopardise their career prospects, or there is fear that diagnosis of illness could ‘leak’. These fears would not necessarily be realised, but this lack of trust could colour the medical encounter. It is becoming acceptable within the British Armed Forces that decisions on health matters are left to the servicemen. In the deployment to the Iraq War for example, servicemen were allowed to decide whether they would accept vaccination against anthrax.

Lack of trust is not the only problem with this type of screening. Goldberg and colleagues validated in civilian populations the GHQ-12 against the Composite Diagnostic Interview (CIDI-PC) in 15 cities in the world.7 The PPVT ranged from 40.6–71.4%; the bottom end of this range was not dissimilar to our results, but their sensitivity was around four higher. Goldberg and Bridge also compared the GHQ and GPs in diagnosing psychiatric illness against two research diagnostic systems and found slightly higher PPVIs in comparison to our study, but the sensitivity was low.4 Kroenke and colleagues6 assessed symptoms count threshold using the PRIME-MD to predict multisomatoform disorder and found that when the threshold was less than 10 symptoms the PPVT was well below the 47% in our study, but their NPVTs were higher. These examples illustrate that even when a standardised gold standard is used the PPVTs are not very high.

If anything it was surprising that the MOs in our study had a similar PPVT, albeit in the lower end of the spectrum, as they are a heterogeneous group in terms of training, or interpretation of the meaning of ‘needing medical help’. Peveler and colleagues, comparing a self-report screening questionnaire and clinical opinion, found that in a large proportion of patients the doctor missed the diagnosis of unexplained physical symptoms or a mood problem.15 It is unclear from that study whether doctors are missing the diagnosis or reinterpreting findings using a different threshold of illness as the same authors indicate that primary care doctors identified the most serious cases, a finding that was not evident in our study.

Our study included several health dimensions so our results may represent an overall assessment of health. In our study there was little difference between the PPVT for multiple symptoms, alcohol behaviour and GHQ, and change of threshold had little impact. The use of the MO as the gold standard was appropriate because this specialist would decide serviceman’s management and course of action in an implemented programme. There is evidence that doctors in primary care are good at distinguishing between organic and non-organic conditions37 and it is probable that they would not perform worse in distinguishing between illness and lack of it. We do not know whether MOs would be different to other primary care doctors. It is possible that our screening detects short duration health problems as it has been shown that approximately 75% of patients improved within a fortnight of their clinical visit.14,19 As the median time between receipt of the screening questionnaire and the consultation was around four months, many ‘screen-positives’ may have changed status over the period. Another possibility would be that servicemen might have exaggerated reports of health in the questionnaire.20 This possibility cannot be discounted in our study, but it is suggestive that the percentage of servicemen needing medical help according to the MOs is similar to the percentage of ‘screen-positives’. Thus if some of the servicemen exaggerated symptoms, others may have under-reported their symptoms.

An important issue to consider is the self-perception of health of a person who is a ‘screen-positive’. It is known that those with an alcohol intake problem do not always perceive themselves as ill or needing medical help.21 It has also been reported that patients who gave a psychological [AQ3] explanation of their symptoms were more likely to consult than those who gave normalising or somatic explanations.22 The decision of an individual to consult a doctor may be related also to the perception of illness severity.23 Only those who believe that the symptoms are serious enough may disclose their symptoms in a medical encounter.

We should also consider the provider perspective, as MOs may not mechanically accept the results of a health
questionnaire. It has been shown that GPs are more likely to initiate treatment for patients whom they themselves have diagnosed as depressed.24,25 There are also doctors who are reluctant to accept test scores as a reasonable procedure for making diagnosis.26

In conclusion the validity of a prospective screening programme on physical and psychological health based on a military primary care setting is unsatisfactory. The value of questionnaires may be reinterpreted or not taken into account by MOS, the course of the possible condition is heterogeneous and the servicemen may be unwilling to confide their symptoms to an MO on referral.

ACKNOWLEDGEMENTS
We are indebted to the administrative services in the Armed Forces, the Defence Analytical Service Agency and the British Forces Post Office at Mill Hill, staff in the medical centres and, especially, the servicemen. This study was funded by the Defence Science and Technology Laboratory (Dstl), part of the MoD. Those who funded the study had no input into the data analysis, results presented or their interpretation. A copy of the manuscript was sent to the last Dstl officer and the chairman of the BSNG for comments before submission

Author’s affiliations
Roberto J Rona, Department of Public Health Sciences, Guy’s, King’s and St Thomas’ School Medicine
Richard Hooper, Department of Public Health Sciences, Guy’s, King’s and St Thomas’ School Medicine
Margaret Jones, Department of Public Health Sciences, Guy’s, King’s and St Thomas’ School Medicine
Claire French, Department of Public Health Sciences, Guy’s, King’s and St Thomas’ School Medicine
Simon Wessely, Academic Department of Psychological Medicine, Institute of Psychiatry [AQ4]

REFERENCES
3 Director-General Defence Health Service (Australia). Mental health support to operationally deployed forces. Health Bulletin No 11/2003.

AQ1: define SF-36
AQ2: Is sentence structure OK?
AQ3: Psychological OK?
AQ4: Please provide author job titles e.g. Professor