Screening for vulnerability to psychological disorders in the military: an historical survey

E Jones, K C Hyams, S Wessely

Methods: An extensive literature review was conducted by hand-searching leading medical and psychological journals relating to World Wars I and II. Recent publications were surveyed electronically and UK archives investigated for British applications.

Results: Despite the optimism shown in World War I and the concerted efforts of World War II, follow-up studies showed that screening programs did not succeed in reducing the incidence of psychological casualties. Furthermore, they had a counter-productive effect on manpower, often rejecting men who would have made good soldiers. Continued experimentation with screening methods for psychiatric vulnerability failed to yield convincing results during the post-war period.

Conclusions: Although well-measured variables, such as intelligence, have been shown to predict success in training and aptitude, no instrument has yet been identified which can accurately assess psychological vulnerability. Previous attempts have failed because of false-positives, false-negatives and reluctance in the target population because of stigma. Early findings suggest that psychological surveillance, if not screening, may yield valuable results when applied to military populations exposed to stress.

Screening for any disorder has an intuitive appeal, and if disorders can be identified before they have become apparent the rewards are even greater. Psychiatric disorders, once established, are often difficult to treat and often have a major impact on the person and society. Hence, numerous attempts have been made to screen not just for established psychiatric disorders but for vulnerability to those disorders. Few organizations have been more concerned with these issues than the military. Recruits who fail in training or leave the service early are a major drain on resources. Servicemen who develop psychiatric disorders often function poorly, while soldiers who develop psychological disorders after combat are rendered ineffective for future service and are likely to qualify for financial resources. Servicemen who develop psychiatric disorders can be identified before they have become apparent the rewards are even greater. Psychiatric disorders, once established, are often difficult to treat and often have a major impact on the person and society. Hence, numerous attempts have been made to screen not just for established psychiatric disorders but for vulnerability to those disorders. Few organizations have been more concerned with these issues than the military. Recruits who fail in training or leave the service early are a major drain on resources. Servicemen who develop psychiatric disorders often function poorly, while soldiers who develop psychological disorders after combat are rendered ineffective for future service and are likely to qualify for financial resources.
Various instruments were designed to assist officer selection and rating, together with assignment to particular trades. This represented the first attempt to predict the best soldiers rather than identify those most likely to suffer from a psychological disorder. A paper and pencil survey, correlated with the standardized Stanford-Binet intelligent test, was developed for the military. In 1917, Verkes, with the help of Lewis Terman, David Wechsler and others, developed multiple-choice Army Alpha and non-verbal Beta tests to measure both literate and illiterate subjects.6

In November 1917, Colonel Henry Shaw reported favourably to the Surgeon General, recommending “that the [screening] scheme be extended to include all enlisted and drafted men and all newly appointed officers, provided competent psychologists can be found to take charge.”3 In the same year, Robert Woodworth devised the personal data sheet, the first group personality test and forerunner of the current Minnesota Multiple Personality Inventory; and this was used briefly to screen recruits. However, a cable sent from General Pershing in July 1918 implied that the effectiveness of these instruments was qualified. “Prevalence of mental disorders in replacement troops recently received was considered the first attempt to predict the best soldiers and, in June 1941, a Directorate of psychiatrists concluded that about 4% of all intakes were psychopathic and allied mental disorders, 8”

During World War I, 72,000 (2%) US recruits were rejected on neuropsychiatric grounds either at boards or in training.7 Draft boards screened out 15.1 men per 1000 registrants for psychopathy and allied mental disorders,8 while a further 0.5% were discharged at their first military station. Latterly, Ginzberg et al. calculated that 68,000 (1.4%) recruits were rejected at enlistment, and a further 35,000 (0.9%) were discharged from the armed forces with neuropsychiatric disorders.9

In February 1919, it was calculated that only 4039 (0.2%) servicemen had been invalidated from the American Expeditionary Force in France with psychological disorders. This was considered a success, given that 3181 soldiers had been evacuated to the United States.6 During World War I, 72,000 (2%) US recruits were rejected on neuropsychiatric grounds either at boards or in training.7

Screening: World War II

The high watermark of screening by the military for vulnerability to psychological disorders came in World War II. The outbreak of hostilities was greeted with warnings about the past, references being made to the immense cost incurred by the post-World War I epidemic of shell shock and other post-combat disorders; their principal weapons were strict regulations about diagnostic terms, restrictive pension criteria and strategies to prevent premature discharge from the forces.10 The belief that manpower resources were not being used effectively provided the main impetus for screening. The increasingly egalitarian stance adopted by some military psychiatrists and psychologists was critical of traditional, class-based selection methods. In April 1940, Lt Colonel Ronald Hargreaves used Penrose-Raven Progressive Matrices, a pre-war test designed to measure innate intelligence, to screen unsuitable recruits.11 As a trial, matrices were given to servicemen diagnosed with effort syndrome undergoing treatment in the special treatment unit at Mill Hill Hospital. The results purported to show that “neurotic men have less consistent test scores [over time] than normal people” and that the scores of the former improved with treatment.12 Army psychiatrists concluded that about 4% of all intakes were unsuitable for combatant units.

Considerable effort was put into instruments designed to identify types of soldier and, in June 1941, a Directorate of
Selection Personnel was established for the British Army. In the following year, new recruits were enlisted into the General Service Corps and posted to Primary Training Centres where they completed a battery of intelligence and aptitude tests. Each serviceman was interviewed by a personnel selection officer, who made recommendations about the man's deployment. Recruits with low test scores, or who were regarded as unstable, were referred to military psychiatrists who could recommend special employment or discharge. Hence, the primary aim of this system was to ensure that servicemen were placed in suitable trades or occupations rather than the identification of potential psychiatric casualties. With more modest aims, the British army rejected far fewer recruits (1.4%) than the US Army (7.2%) during World War II.41

The limited screens employed by the British reflected a widespread mistrust of psychological testing and the reduced levels of choice offered by a much smaller population.

Outcomes of screening: United States

By spring 1943, it had become obvious from the large numbers of psychiatric casualties that the screening programme had failed in its preventive role. Admissions for neuropsychiatric disorders had risen to 20,000 per month in the US alone and reached a peak of 31,000 in August. During the last month of the Buna-Gona campaign, the neuropsychiatric admission rate was 60-70 per 1000 troops for the entire Southwest Pacific area – four times that for the American Expeditionary Force during World War I.29,30 Whereas still, the programme was having an adverse effect on the war effort. Selection programmes rejected too many people, adding to the general manpower problems. Official directives documented the disillusionment, which culminated in the War Department Technical Bulletin (TR MED 33) issued on 21 April 1944. This stated that rejection for neuropsychiatric conditions was also greater. Hence, despite a tough rejection policy, "large numbers of men with emotional difficulties and neurotic trends could be of service".

According to official statistics prepared by the US Office of the Surgeon General, 7.2% of those registered for the draft were turned down for neuropsychiatric reasons and a further 4.3% because of low intelligence or educational defects and neurotic trends could be of service". Official directives documented the disillusionment, which culminated in the War Department Technical Bulletin (TR MED 33) issued on 21 April 1944. This stated that rejection for neuropsychiatric conditions was also greater. Hence, despite a tough rejection policy, "large numbers of men with emotional difficulties and neurotic trends could be of service".

Outcomes of screening: United Kingdom

In the UK, the Director of Manpower Planning at the War Office argued for the extension of aptitude screening in the post-war period so that it could be employed during National Service: "yet the need to make economies and an enduring suspension of these methods brought an end to the General Service Corps in 1948. Henceforth, recruits were admitted directly to corps or regiments after a preliminary selection test, the remainder of their assessment being completed during training. The primacy of the regimental system was restored and the part played by psychological testing in allocating servicemen to different units downgraded. Nevertheless, the principle that these techniques could be used for personnel selection officer and trade selection had been established and remains so to this day.

Why did screening fail in World War II?

At the same time that doctors and officials had been calling for a screening programme, voices were also heard warning against over-optimistic claims. Based on his experience treating World War 1 veterans, Kardiner wrote "I should hesitate to offer any criteria that can be used to predict that a given candidate will have a traumatic neurasthenia." Aita concluded that the screening techniques of 1941 had little value because they relied on past behaviour. Predicting how these men would perform in circumstances quite different from anything they had experienced hitherto involved guesswork.

In the immediate post-war period a number of studies were conducted to discover why the screening programmes had failed. A broad-ranging personnel selection inventory was conducted by Ellis and Conrad concluded that "except in samples containing an unusually large proportion of psychiatrically 'positive' cases, the number of cases falsely classified as positive by the inventory generally exceeds, by a great deal, the number correctly classified as positive." Egan and colleagues followed up 2054 men rejected by the Selective Service System on psychiatric grounds, but later inducted into the army – an elegant natural experiment. Although their sample was not entirely random, the study showed that only 18% had subsequently been discharged from the forces on psychiatric grounds. Of the rest, 82% had given "satisfactory duty" as compared with 94% of all enlisted personnel. They estimated that 1,992,950 men had been unnecessarily rejected for military service on psychiatric grounds during World War II. The variables used to reject those thought vulnerable to later breakdown did, indeed, have low predictive power. Brill and Beebe compared the rates of psychiatric admission for World War II servicemen who had been assessed as being at risk with those considered well-adjusted. In those without such a history, 2% had been admitted to a psychiatric facility compared with 29% of the marginal or predisposed group. If allowance was made for those discharged with a diagnosis of personality disorder but never formally admitted for psychiatric care, the total rose to 35%. Although this research confirmed that the likelihood of admission was substantially higher in the identified population, it also demonstrated the impracticalities of screening for such predispositions. At least 65% of an at-risk population rendered satisfactory service. Had the examiners rejected everyone from this marginal group at induction, the army would have been deprived of one million men who never broke down.

As regards the related question of how those diagnosed with psychiatric disorders actually performed in combat, Plesset followed up 138 soldiers who in training had shown...
“sufficient adjustment difficulty to necessitate psychiatric attention”. After 30 days of combat, 137 remained on active duty, and one had received a gallantry medal. By the end of the war, 120 remained on duty and eight had been awarded Bronze Stars. A further study investigated the performance of 395 men diagnosed during training with mild psychological disorders. Subjected to combat during the Ardennes campaign, only nine became psychiatric casualties. Eleven of the 120 servicemen who were “merely transitory”. The authors appeared to pay little attention to the most obvious explanation – that the instruments performed badly – though they did conclude that “psychiatric screening has . . . limited value”. Huffman suggested that a factor in the low incidence of psychiatric cases during the Vietnam War was the effectiveness of screening. However, the US Army had not employed an organised screening programme for inductees to assess psychological vulnerability during the conflict. Glass and his colleagues had raised sufficient doubt about the accuracy of psychiatric predictions of both military effectiveness and the likelihood of developing psychological disorders. Nevertheless, the military encountered recruitment difficulties throughout what became an unpopular war, so that there was little scope to reject both volunteers and conscripts. Nevertheless, the military had some success with the more limited goal of attempting to identify recruits who would complete basic training. Elements of an 82-item questionnaire dealing with personal adjustments and attitudes, which was given to 9194 airforce trainees, were said to predict outcome with “uncanny accuracy.” Subsequently, the Health Opinion Survey (HOS) applied to 1462 recruits at their induction and to the 1167 who completed their training gave only broad statistical risk predictions. The raw data showed that the scale had unacceptably low specificity and sensitivity for routine use.

Korean War

The US army, to its credit, put into practice some of the lessons learned in the aftermath of World War I, namely that screening was more effective if variables could be measured accurately. A retrospective study of a rapid screening procedure applied in 1943–44 to assess intellectual and emotional deficiencies in the US Marine Corps had reportedly identified 78% of those subsequently discharged, though it also generated 11% false-positives. Nevertheless, it was also shown that the judgements of drill instructors in weeding out unsuitable men proved no less efficient than the screen instruments, suggesting that these were observable disorders rather than potential vulnerabilities. In the light of similar studies, the emphasis placed on intelligence was increased, while less attention was paid to the detection of psychological vulnerability, and a more liberal policy was adopted towards so-called “borderline cases.” As a result of more modest aims, the rate of neuro-psychiatric rejections fell from the 7.2% recorded during World War I to 0.1% during the Korean conflict. However, those excluded on grounds of mental deficiency rose from 4.3% to 13.4%, largely because of the emphasis given to intelligence testing. In the aftermath of the war, Voth argued that “exception is taken to the premise that brief psychiatric screening is of value in predicting possible psychiatric casualties, except in certain obvious cases.” Glass and colleagues retrospectively examined a random sample of 505 inductees, who had entered training in August 1951 and hence already passed through the psychiatric screening process. An analysis of service records and evaluations by unit commanders showed that prediction of military effectiveness was unreliable. For combat and combat-support troops deployed to Korea, forecasts of below-average or poor performance had been unduly pessimistic. Troops sent overseas to non-combat theatres also performed far better than predicted. Greater accuracy was obtained with servicemen assigned non-combat roles in the United States, screening having identified 88.8% of those who were unable to continue. Yet most of these servicemen had broken down during or shortly after training, and were soldiers who had experienced difficulties adjusting from civilian to military life. Furthermore, authorities concluded that overt signs of psychiatric abnormality could only forecast military effectiveness over a relatively brief period.

Post-1953

The search for an accurate screening instrument did not cease after the Korean War. The Fort Ord Inventory (FOI) was designed to identify affective disturbances that would impair military service. Tests on 15,000 recruits confirmed the design of four scales, which differentiated between those with leadership potential and those with poor adjustment qualities: delinquency, neurosis, lacks hope (malingerers) and femininity. Although the authors concluded that they could detect “the major types of emotional abnormalities”, the FOI was not designed to predict how soldiers might perform in combat. A follow-up study of 134 naval recruits enlisted between 1960 and 1961, who had been rejected on psychological grounds but had been purposely allowed to graduate from training, found that two years later 97 (72.4%) were still on active duty. It was hypothesised that most had experienced “emotional growth” during their service and that the disturbances picked up on screening were “merely transitory”. The authors appeared to pay little attention to the most obvious explanation – that the instruments performed badly – though they did conclude that “psychiatric screening has . . . limited value.”
Another study, which used a complex battery of physical and psychological measures to predict who would complete basic training, gave borderline results for female recruits but proved unreliable for males.55 Some successes were recorded, though the variables that could be measured most accurately were straightforward, largely concerned with age, education or reason for enlistment,66 reiterating one of the basic requirements for screening that variables be measured accurately. The US Air Force persisted with personality assessments to assess those in training, but only achieved a predictive power of 50%.67 Not deterred, a research programme at Wilford Hall Medical Center, Texas, designed the Air Force Medical Evaluation Test Programme (AFMET) to identify trainees with significant psychological problems.68 Although AFMET has plausibly saved valuable resources through early discharges, the system did encounter problems, including “excessive misidentifications”.69 A further attempt to assess outcomes in US Air Force trainees suggested promising results, though the authors conceded that further research was needed to improve accuracy.70 By 1990, disillusionment with personality testing was widespread, not just in the military.64

### Post-traumatic stress disorder (PTSD)

The official recognition of PTSD in 1980 opened a new chapter in the history of psychological screening as investigators attempted to identify risk factors for this high-profile diagnosis. A meta-analysis by Brewin et al. of 45 retrospective and prospective studies used statistical techniques to increase the predictive power of each investigation and overcome some of the problems associated with small sample sizes.71 Overall, the results showed that no single variable was a particularly useful predictor of developing PTSD. The best indicator, though by no means exceptionally strong, was the intensity of the trauma itself. This was not a new finding, having been identified in World War II studies of combat fatigue. By definition, such a factor cannot be used in recruitment or pre-deployment screening.

Associations with PTSD were noted for other variables, which could be measured before combat, and included gender, age, social class, intelligence, education, family, personal histories of psychiatric disorder and, more controversially, childhood abuse. None of these associations were statistically significant (because meta-analytic techniques had increased the sample size) but actually were very small. Little heterogeneity was found amongst the risk factors identified, and only previous psychiatric history emerged as a uniform variable. As a result, Brewin concluded that “attempts to identify a common set of pre-trauma predictors of PTSD that will be equally valid across different traumatized groups are premature”.72

### Current policy: Bosnia and Kosovo

In February 1996 all US military personnel deployed to Bosnia for more than 30 days were required to complete a mental health screen as they returned to their home station. Respondents who exceeded pre-established, cut-off criteria on the mental health screen as they returned to their home station. Respondents who exceeded pre-established, cut-off criteria on the mental health screen as they returned to their home station. Respondents who exceeded pre-established, cut-off criteria on the mental health screen as they returned to their home station. Respondents who exceeded pre-established, cut-off criteria on the mental health screen as they returned to their home station. Respondents who exceeded pre-established, cut-off criteria on the mental health screen as they returned to their home station. Although some successes were recorded, though the variables that could be measured most accurately were straightforward, largely concerned with age, education or reason for enlistment,66 reiterating one of the basic requirements for screening that variables be measured accurately. The US Air Force persisted with personality assessments to assess those in training, but only achieved a predictive power of 50%.67 Not deterred, a research programme at Wilford Hall Medical Center, Texas, designed the Air Force Medical Evaluation Test Programme (AFMET) to identify trainees with significant psychological problems.68 Although AFMET has plausibly saved valuable resources through early discharges, the system did encounter problems, including “excessive misidentifications”.69 A further attempt to assess outcomes in US Air Force trainees suggested promising results, though the authors conceded that further research was needed to improve accuracy.70 By 1990, disillusionment with personality testing was widespread, not just in the military.64

### DISCUSSION

Although screening for psychological vulnerability to breakdown under stress remains a most desirable goal, this review suggests that its achievement has been elusive. By contrast, screening has been shown to be effective for certain well-measured variables, such as low intelligence and psychosis, which are powerful predictors of failure in military service. Instruments to assess aptitude have been progressively refined and are used by all the main military powers to assist officer selection and trade deployment. Screening has also been shown to identify accurately those who are unable to complete their basic training, saving valuable resources.

However, screening becomes more problematic when applied to prediction of breakdown under stress. Some key variables, such as leadership, morale, intensity and result of battle, preparedness, which contribute to functioning, cannot be known at induction. The imprecision of current measurement also presents an obstacle to identification. Our knowledge of what makes people subject to PTSD remains relatively crude, including past psychiatric history, social class, family history and childhood abuse. None of these have sufficient explanatory power to justify a screening programme.73

While no variable or combination of variables has yet been found that will accurately identify vulnerability, some historical data suggest that significant predictors may exist. In their retrospective study, Brill and Breeie discovered that “the vast majority of army admissions in 1942 and 1943 involved men with clear-cut, pre-existing emotional disorders”, while in 1944 and 1945, when combat was the dominant factor, 50% were found to have a history of psychological problems.74 Although Atta concluded that the screening techniques of 1941 were too imprecise to justify their continued use, these instruments had succeeded in identifying psychiatric casualty rates that were three times higher than in the controls. The problem was that 60% of those identified as borderline proved to be efficient soldiers, undermining any gains that would have accrued in rejecting potential breakdowns. Hence, even if our current knowledge is inadequate, it is possible that further research into psychosocial and possibly genetic factors that convey vulnerability may contribute to the design of an effective screening instrument.

There still remains the problem of how to avoid rejecting those recruits with a history of psychological illness who in practice would have proved to be good soldiers. The disastrous screening programme of the early years of World
Simon Weasely, Consultant, Office of Public Health and Environmental Hazards, Department of Veterans Affairs, Washington DC, USA, and Professor of Psychological Medicine, Guy’s, King’s and St Thomas’s School of Medicine, London, UK.

REFERENCES

6 Mayfield Campbell C. Selective service and psychiatric issues. JAMA 1941;118:1884.
15 Lester A. The utilization of psychiatry in the armed forces. Psychiatry 1942;2:2.
18 Davidson HA. Mental hygiene in our armed forces. MH Surgeon 1940;84:480.
Screening for vulnerability to psychological disorders

37 Kardiner A. The neuroses of war. War Med 1941;1:221.
38 Arrn JA. Neuropsychologic and psychiatric examination during military mobilization. War Med 1941;3:775.
44 Sharp WL. Fate of 395 mild neuropsychiatric cases salvaged from training period and taken into combat. Am J Psychiatry 1950;106:601-7.