Running head: MILITARY MENTAL HEALTH SCREENING

The validity of military screening for mental health problems: Diagnostic accuracy of the PCL, K10 and AUDIT scales in an entire military population

Dr. Amelia K. Searle[1], Dr. Miranda Van Hooff[1], Prof. Alexander C. McFarlane[1], Mr. Christopher E. Davies[2], Dr. A. Kate Fairweather-Schmidt[3], Dr. Stephanie E. Hodson[4], Ms. Helen Benassi[5], & Ms. Nicole Steele[5]

[1]Centre for Traumatic Stress Studies, the University of Adelaide, South Australia
[2]Data Management and Analysis Centre, Discipline of Public Health, the University of Adelaide, South Australia
[3]School of Psychology, Flinders University, South Australia
[4]Department of Veterans' Affairs, Canberra ACT
[5]Mental Health, Psychology and Rehabilitation Branch, Joint Health Command, Department of Defence

Correspondence to Dr Amelia K Searle: amelia.searle@adelaide.edu.au; Level 2/122 Frome Street, Adelaide, South Australia, AUSTRALIA 5000. Ph: +61 8 8313 5200 F: +61 8 8313 5368

Abstract

Depression, alcohol use disorders and post-traumatic stress disorder (PTSD) are serious issues among military personnel due to their impact on operational capability and individual wellbeing. Several military forces screen for these disorders using scales including the Kessler Psychological Distress Scale (K10), Alcohol Use Disorders Identification Test (AUDIT), and Post-traumatic Stress Disorder Checklist (PCL). However, it is unknown whether established cut-offs apply to military populations. This study is the first to test the diagnostic accuracy of these three scales in a population-based military cohort.

A large sample of currently-serving Australian Defence Force (ADF) Navy, Army and Air Force personnel (*n* = 24481) completed the K10, AUDIT and PCL-C. Then, a stratified subsample (*n* = 1798) completed a structured diagnostic interview detecting 30-day disorder. Data were weighted to represent the ADF population (*n* = 50049).

ROC analyses suggested all three scales had acceptable sensitivity and specificity, with areas under the curve from .75 to .93. AUDIT and K10 screening cut-offs closely paralleled established cut-offs, whereas the PCL-C screening cut-off resembled that recommended for US military personnel.

These self-report scales represent a cost-effective and clinically-useful means of screening personnel for disorder. Military populations may need lower cut-offs than civilians to screen for PTSD.

Key words: military; sensitivity and specificity; K10; PCL; AUDIT

Experiencing deployment-related trauma like direct combat and witnessing atrocities (rather than simply having deployed) is significantly associated with subsequent disorder in military personnel, including PTSD, depression and alcohol use disorder (Fear et al., 2010; Hoge, Auchterlonie, & Milliken, 2006; Hoge et al., 2004; Iversen et al., 2008; Sareen et al., 2007). Although mental disorders have negative repercussions for military productivity and personal wellbeing (e.g., Erbes, Meis, Polusny, & Compton, 2011; Hoge et al., 2006; Hoge et al., 2002; Hoge et al., 2005; Rona et al., 2009), they often go untreated, seen in the relatively low service use rates among affected personnel (Hoge et al., 2004; Kim, Thomas, Wilk, Castro, & Hoge, 2010; Sareen et al., 2007). Thus, it is imperative that military forces are better able to identify mental disorders. To achieve this, more information is needed regarding the accuracy of military screening instruments; particularly, whether they can sensitively detect mental disorders among personnel, at which cut-offs they function optimally, and where further identification resources should be directed if needed. This study examines the diagnostic accuracy of several commonly-used scales to screen for mental disorders in a large military sample.

Screening for Mental Disorders in the Military

Several nations (e.g., Canada, the US, New Zealand, and Australia) conduct mental health screening for personnel returning from deployment, to identify those most likely to benefit from an intensive diagnostic interview, target those at-risk for education and prevention, and refer disordered personnel to services (Rona, Hyams, & Wessely, 2005; Steele & Twomey, 2008). Various self-completed questionnaires assess depressive and post-traumatic stress symptoms and alcohol use; these are often accompanied by brief semi-structured interviews to contextualise questionnaire responses and provide brief intervention (Steele & Twomey, 2008).

However, the diagnostic accuracy of these screens has not been clearly established in serving personnel; thus, the effectiveness of military screening programs remains unknown (Dunt, 2009). It is vital that military forces can be confident their screening measures are identifying the correct people, and that benefits of screening outweigh the costs (Rona et al., 2005). Thus, careful implementation

must occur alongside thorough testing (Bliese et al., 2008; Dunt, 2009; French, Rona, Jones, & Wessely, 2004; Rona, Hooper, Jones, French, & Wessely, 2004; Rona et al., 2005; Wright et al., 2005).

In particular, it cannot be assumed that scale cut-offs derived from civilian samples apply in military settings: personnel may require different cut-offs for reasons like regular scale completion (which may lead to habitual responding, or learning the response pattern needed to avoid follow-up) and an ethos of 'fighting through' distress. Additionally, a significant proportion of personnel perceive that seeking help for problems will result in social stigma, including experiencing career harm, and being stopped from deploying (French et al., 2004; Gould et al., 2010; Hoge et al., 2004; Iversen et al., 2011; McFarlane et al., 2011; Sareen et al., 2007). Relatedly, a greater proportion of personnel screen positive for disorders within de-identified research than when results are identifiable within post-deployment screening, with many personnel reporting feeling reluctant to disclose problems during post-deployment screening (Warner et al., 2011).

Suboptimal cut-offs may result in poor sensitivity (the proportion of disordered personnel who are correctly identified) and/or specificity (the proportion of non-disordered personnel who are correctly identified), which both have negative repercussions for personnel. Specifically, if cut-offs are too high and sensitivity too low, a significant proportion of psychologically vulnerable personnel may go undetected and sent into combat. Alternatively, low cut-offs accompanied by low specificity may subject disorder-free personnel to stigmatising attitudes, and increase the workload of mental health service providers to unmanageable levels. Thus, it is critical to determine optimal cut-offs for military personnel.

The Australian Defence Force (ADF) Mental Health Screen

The ADF uses the Return to Australia Psychological Screen (RtAPS) upon departing the area of operations, and the Post-Operational Psychological Screen (POPS) three to six months after returning (Department of Defence, 2008; Dunt, 2009; Steele & Goodman, 2006). These instruments guide a brief universal semi-structured interview conducted by a mental health professional. Three self-report scales are used: the Alcohol Use Disorders Identification Test (AUDIT) assesses alcohol consumption (Babor,

Higgins-Biddle, Saunders, & Monteiro, 2001); the Post-Traumatic Stress Disorder Checklist – Civilian version (PCL-C) assesses post-traumatic stress symptoms (Weathers, Litz, Herman, Huska, & Keane, 1993); and the Kessler Psychological Distress Scale assesses general psychological distress, meaning it can detect symptoms shared between several common mental disorders, like anxiety and depressive symptoms (K10: Kessler et al., 2002) (Department of Defence, 2009; Dunt, 2009). These scales are also used within Defence primary health care and periodic health examinations. The AUDIT and PCL have also been used to screen military personnel internationally (Rona, Jones, French, Hooper, & Wessely, 2004; Steele & Twomey, 2008; Wright et al., 2005).

The ADF selected these scales given their validation and use in international military studies and Australian community/veteran populations (Department of Defence, 2009; Nicholson, 2006). However, their diagnostic accuracy - including optimal screening cut-offs - has not been adequately confirmed within currently-serving military samples.

Diagnostic accuracy of ADF screening scales.

The PCL (Weathers et al., 1993) was developed in Vietnam combat veterans. Although it has shown good overall diagnostic accuracy in primary care and veteran samples (see McDonald & Calhoun, 2010), there is evidence that the established screening cut-off score of 50 is not optimal in all populations/settings, with the optimal[1] cut-off varying from 30 to 60 (McDonald & Calhoun, 2010). This between-sample variability highlights the need to obtain validation evidence for the population to be screened. In the only study of currently-serving military personnel (Bliese et al., 2008), the cut-off of 50 resulted in near-perfect specificity (.98) but poor sensitivity (.24), suggesting that this cut-off would be best used for estimating true population disorder prevalence within epidemiological research. Alternatively, a cut-off of 30 produced high specificity (.88) and sensitivity (.78). As data came from mandatory post-deployment screening, this lower cut-off may have reflected under-reporting due to fear of social stigma. While Bliese and colleagues' (2008) results highlight that military personnel may

[1] As there is no one indicator of optimal performance, optimal cut-offs have been defined in various ways, including those that (1) maximise overall efficiency, (2) maximise the sum of sensitivity and specificity, and (3) balance sensitivity and specificity (see McDonald & Calhoun, 2010).

require different cut-offs, it is unclear whether their results, developed from a relatively small sample (*n* = 352) of US Army personnel recently-returned from combat deployment in Iraq, apply to broader military populations, including personnel returned from any operational deployment, as well as personnel who have never deployed.

The AUDIT (Babor et al., 2001) shows high sensitivity (around the high 80s) in primary care and epidemiological samples using the recommended screening cut-off of 8, with slightly lower though acceptable specificity (Degenhardt, Conigrave, Wutzke, & Saunders, 2001; Reinert & Allen, 2002). However, in seemingly the only diagnostic accuracy study in currently-serving military personnel, a slightly higher optimal cut-off of 10 was found for detecting 12-month DSM-IV alcohol disorder in male Australian Navy veterans (sensitivity = .85, specificity = .77) (McKenzie et al., 2006).

The K10 (Kessler et al., 2002) shows high levels of overall diagnostic accuracy in numerous population-level studies, with areas under the curve from .80 to .96 (Andrews & Slade, 2001; Furukawa, Kessler, Slade, & Andrews, 2003; Kessler et al., 2002; Kessler et al., 2003; Oakley Browne et al., 2010). However, its diagnostic accuracy has not been examined in military populations. The slightly shorter K6 demonstrated high specificity in US military personnel, but sensitivity that was too low for a military screen (Wright et al., 2007). Importantly, no information regarding sensitivity and specificity of particular cut-offs is available. Two sets of cut-offs are used in Australia: (1) national surveys use scores ≥16 and ≥ 21 to indicate moderate and high distress, respectively, and (2) primary care settings use scores ≥ 20 and ≥ 25 to indicate mild and moderate disorder, respectively (Australian Bureau of Statistics, 2003).

In sum, due to the few military diagnostic accuracy studies (among small and specific samples), and the possibility that military personnel may need distinct cut-offs, there is a great need to establish optimal screening cut-offs in actively-serving military populations.

The nature of epidemiological cut-offs

These screening scales are also widely used to estimate population prevalence in military epidemiological studies, given it is impractical to administer diagnostic interviews to large samples (e.g., Fear et al., 2010; Riddle et al., 2007). However, optimal screening cut-offs (identified through sensitivity

and specificity indices) are often quite different from optimal epidemiological cut-offs, with the distinction between them often misunderstood (McDonald & Calhoun, 2010). For screening, a relatively low cut-off is preferable as it is generally desirable for few true positives to be missed. However, for epidemiological purposes, it is important that the number of false classifications is minimal and those screening positive actually have the disorder, meaning a relatively high cut-off is preferable (McDonald & Calhoun, 2010). Furthermore, the number of incorrect classifications (and thus the optimal epidemiological cut-off) is greatly impacted by population disorder prevalence, so that when prevalence is relatively low (as often seen in military populations e.g., Sareen et al., 2007), screening cut-offs will tend to over-estimate prevalence even when sensitivity and specificity are high (McDonald & Calhoun, 2010; Terhakopian Sinaii, Engel, Schnurr, & Hoge, 2008). In these cases, higher epidemiological cut-offs are needed, even though they result in lower sensitivity. As these two types of cut-offs are distinct and not interchangeable, it is important to establish both within the same population.

The aim of this study was to examine the diagnostic accuracy of these three screening scales (the AUDIT, PCL-C, and K10) against 'gold standard' structured diagnostic interviews. This is the first study to assess these scales in a large and representative sample of actively-serving military personnel. While our primary aim within the context of military screening was to establish optimal *screening* cut-offs, a secondary aim was to identify optimal *epidemiological* cut-offs, after establishing the prevalence of disorder in this military population.

Method

Participants

Participants came from the 2010 ADF Mental Health Prevalence and Wellbeing Study (MHPWS: McFarlane et al., 2011), which measured the prevalence of mental disorders in a representative sample of currently-serving ADF personnel. Detailed methodology is described elsewhere (McFarlane et al., 2011).

A two-phase assessment was used. First, all currently-serving ADF personnel in the Navy, Army, and Air Force (as at 11th December 2009) but excluding trainees and reservists were considered eligible and contacted for Phase 1 participation: this was 50049 personnel. Of these, 24481 (49% of the ADF population) agreed to participate and completed self-report questionnaires. Among the remaining 25568 eligible personnel who did not participate, 76% never responded (and some may not have even read our correspondence); 17% actively declined; and 7% consented, but never completed a questionnaire. Second, a stratified sub-sample of 3688 (15% of the Phase 1 sample) were sought for Phase 2, of which 1798 (49% response rate) completed a telephone interview. The interview sub-sample pool was stratified by Service, sex (oversampling for females, to ensure sufficient numbers in each Service), and the combination of members' Phase 1 screening scores (oversampling for high scorers, to reduce the possibility of error in prevalence estimates by limiting the number without mental disorders).

Table 1 provides sample demographic characteristics. The sample was predominantly male (76%), and aged, on average, 38.3 years (*SD* = 9.4). The sample comprised members from all Services (40% Army, 39% Air Force, and 21% Navy) and ranks (49% commissioned officers, 36% non-commissioned officers, and 14% other ranks). Compared with the total ADF population, respondents were slightly older and had served for longer, comprised a greater proportion of personnel who were married and in the Army, and a smaller proportion of males, deployed personnel, and personnel in other ranks. This sample was not intended to resemble the ADF population, as females and those with higher screening questionnaire scores were oversampled. Moreover, these observed differences were subsequently used in the population weighting process, so that the estimates generated effectively represented the entire ADF population.

Measures

Screening scales (the index tests).

Alcohol Use Disorders Identification Test (AUDIT).

The AUDIT comprises 10 questions on alcohol consumption, dependence and problems, typically or in the last 12 months. Total scores range from 0 to 40, with higher scores indicating more problematic alcohol consumption. The AUDIT demonstrates high internal consistency, factorial convergent and criterion validity (Allen, Litten, Fertig, & Babor, 1997; Degenhardt et al., 2001; Reinert & Allen, 2002). Internal consistency was good in our sample (alpha = .75). The ADF uses a screening cut-off of 8 (warranting simple advice), with scores above 20 resulting in comprehensive assessment and referral to drug/alcohol services (Department of Defence, 2009).

Post-traumatic Stress Disorder Checklist - civilian version (PCL-C).

The ADF uses the PCL civilian version (PCL-C: Weathers et al., 1993) , as it allows members' ratings to be based on any trauma, not just trauma experienced during military service (Nicholson, 2006). The 17 questions correspond with the DSM-IV PTSD symptomatic criteria. Respondents rate these symptoms in the past month which, once summed, give a total score ranging from 17 to 85, with higher scores indicating higher levels of PTSD symptoms. Overall, the PCL shows high validity and reliability (McDonald & Calhoun, 2010; Wilkins, Lang, & Norman, 2011). Internal consistency was excellent in our sample (alpha = .95). The ADF uses a screening cut-off (indicating the need for psychologist follow-up) of 30, with scores above 50 triggering automatic referral (Department of Defence, 2009; Nicholson, 2006).

Kessler Psychological Distress scale (K10).

As a measure of general psychological distress, the K10 (Kessler et al., 2002) detects symptoms found in several common disorders, including depressive and anxiety symptomatology. Participants rate the 10 questions in reference to the last 4 weeks. Total scores range from 10 to 50[2],

[2] The Australian scoring system is different from the US system, where each response is scored from 0 to 4, and total scores range from 0 – 40 (see http://www.hcp.med.harvard.edu/ncs/k6_scales.php)

with higher scores indicating higher psychological distress. The K10 is widely used in clinical screening and epidemiological research, shows high factorial validity and internal consistency, and performs as well as/better than other relevant questionnaires (Andrews & Slade, 2001; Baillie, 2005; Furukawa et al., 2003; Hides et al., 2007; Kessler et al., 2002; Kessler & Üstün, 2004). Internal consistency was excellent in our sample (alpha = .91). The ADF uses a screening cut-off of 20 (indicating the need for follow-up, and potential referral) (Department of Defence, 2009; McFarlane et al., 2011).

Structured diagnostic interview (the reference standard).

Selected sections of the computerised Composite International Diagnostic Interview 3.0 (CIDI: Kessler & Üstün, 2004) were administered by trained Psychology (Honours) graduates via telephone. The modules administered were depression, mania, panic disorder, specific and social phobia, agoraphobia, generalised anxiety disorder, obsessive-compulsive disorder, PTSD, alcohol use, tobacco, and separation anxiety (although the last two modules were not used here); all other CIDI modules (e.g., psychosis, personality) were not administered. The World Health Organization's International Classification of Diseases system (ICD-10: World Health Organization, 1992) was used to diagnose 30-day anxiety disorder, affective disorder, PTSD, alcohol harmful use[3], and alcohol dependence. The CIDI is widely used in epidemiological surveys, and shows high convergent and predictive validity (Haro et al., 2006).

Procedure

Data collection spanned April 2010 and January 2011. In Phase 1, personnel were contacted by email and mail, to seek participation and distribute study materials. Emails, letters, defence base visits and telephone calls followed-up non-respondents. For Phase 2, the Phase 1 participants who were selected as eligible for the CIDI interview sample (through the abovementioned stratified sampling process) were telephoned and invited to complete a telephone interview. Only those who could be interviewed within 60 days of completing their questionnaire were eligible. At most, 10 phone call

[3] Although the specific alcohol use disorder 'alcohol harmful use' has also been referred to as 'alcohol abuse' (including within the DSM-IV classification system), we refer to 'alcohol harmful use' herein, to be consistent with the WHO ICD-10 classification used in this paper.

attempts were made (as well as two recorded telephone messages). Informed consent was digitally recorded via telephone. Interviewers were blind to participants' screening scores. On average, 42 days (*SD* = 25.3) elapsed between survey and interview completion. Interviews took, on average, 30 minutes for non-symptomatic and 60 minutes for symptomatic personnel.

This study was approved by the Australian Defence Human Research Ethics Committee, the University of Queensland Behavioural and Social Sciences Ethical Review Committee, the Department of Veterans' Affairs Human Research Ethics Committee and the University of Adelaide Human Research Ethics Committee.

Statistical Analyses

Statistical analyses were conducted in SAS version 9.2 and Stata version 11.2. Data were weighted to correct for differential non-response, and obtain prevalence estimates for the entire ADF population. Questionnaire results were weighted by sex, Service, rank and medical employment classification (MEC) status. CIDI results were weighted using the interview selection strata (Service, sex and Phase 1 screening scores). Within each stratum the weight was calculated as the population size divided by the number of stratum respondents. A finite population correction was also applied to adjust variance estimates for the reasonably large sampling fraction within each stratum.

Receiver Operating Characteristic (ROC) analysis evaluated screening scale cut-offs for detecting 30-day ICD-10 disorders (the criterion variables). Diagnostic accuracy was evaluated with respect to: (1) the area under the ROC curve (or AUROC, representing the probability that a randomly selected participant with the specified disorder scores higher than a randomly selected member without the disorder) (2) sensitivity (the probability of accurately detecting those with a specified disorder using the specified cut-off) (3) and specificity (the probability of correctly identifying those who <u>do not</u> have the specified disorder using the specified cut-off), (4) overall diagnostic efficiency (the proportion of the total sample that has been correctly identified), (5) positive predictive value (the proportion of those screening positive who have the disorder), and (6) negative predictive value (the proportion of those

screening negative who do not have the disorder). Weighted estimates of proportions were used to estimate these indices. Jackknife sampling was used for the estimation of AUROC and standard error.

Using ROC analysis, we identified two optimal cut-offs for each scale, corresponding with our primary and secondary aims:

1. The **screening cut-off** maximised the sum of the sensitivity and specificity (the *proportion* of those with and without the disorder that are correctly classified), and is suited to identify personnel who *might* need care.
2. The **epidemiological cut-off** brought the *number* of false positives (incorrect disorder identifications) and false negatives (missed disorder identifications) closest together, counterbalancing these sources of error most accurately. Therefore, this cut-off would give the closest estimate to the true prevalence of 30-day disorder, and is suited to monitor trends.

Epidemiological cut-offs are always higher than screening cut-offs as they aim to identify only those with clinical disorders, whereas screening cut-offs are designed to be more inclusive, given that any false positives may be ruled out following diagnostic interview.

Results

Descriptive statistics are provided in Tables 2 and 3. Military personnel showed low symptomatology according to the screening scales, with mean values towards lower scale limits. The prevalence of 30-day disorder ranged from 9.2% for any anxiety/affective disorder, to 0.2% for alcohol harmful use.

We present abridged diagnostic accuracy tables, which include the optimal screening and epidemiological scores, and one score above and below each cut-off (full tables available upon request).

AUDIT

Table 4 presents optimal AUDIT scores for detecting 30-day ICD-10 alcohol harmful use/dependence.

The AUROC for detecting any 30-day ICD-10 *alcohol harmful use* was 0.87 (95% CI 0.72-0.98), indicating good discriminating value. The optimal *screening cut-off* was 8: using this cut-off, the AUDIT had a sensitivity of 1.00 (95% CI 1.00-1.00), indicating this cut-off would detect 100% of those with alcohol harmful use, and a specificity of 0.75 (95% CI 0.73-0.78), indicating that 75% of those without disorder will score below this cut-off. This cut-off resulted in 12350 false positive and 0 false negative diagnoses. In contrast, the optimal *epidemiological cut-off* was 26, with a specificity of 1.00 (95% CI 1.00-1.00) (but with zero sensitivity). This cut-off resulted in 109 false positive and 118 false negative diagnoses.

The AUROC for detecting any 30-day ICD-10 *alcohol dependence* was 0.93 (95% CI 0.89-0.97), indicating good discriminating value. The optimal screening cut-off was 9: sensitivity was 0.91 (95% CI 0.81-1.00), and specificity was 0.83 (95% CI 0.81-0.85). This cut-off resulted in 8688 false positive and 34 false negative diagnoses. In contrast, the optimal *epidemiological cut-off* was 21, with a specificity of 0.99 (95% CI 0.99-1.00) (but a sensitivity of 0.08 (95% CI 0.01-0.18)). This cut-off resulted in 316 false positive and 347 false negative diagnoses.

The AUROC for detecting 30-day ICD-10 *any alcohol disorder* was 0.91 (95% CI 0.87-0.96), indicating good discriminating value (see Figure 1). The optimal screening cut-off was 8: sensitivity was 0.95 (95% CI 0.89-1.00), and specificity was 0.76 (95% CI 0.73-0.78). This cut-off resulted in 11996 false positive and 23 false negative diagnoses. In contrast, the optimal *epidemiological cut-off* was 20, with a specificity of 0.99 (95% CI 0.99-1.00) (but a sensitivity of 0.19 (95% CI 0.02-0.37)). This cut-off resulted in 343 false positive and 400 false negative diagnoses.

Overall, conservative screening and epidemiological cut-off values of 8 and 20 (respectively) would identify both alcohol harmful use and dependence.

PCL-C

Table 5 presents optimal PCL-C scores for detecting 30-day ICD-10 PTSD, and Figure 1 presents the ROC curve. The AUROC was 0.85 (CI 95% 0.79-0.91), indicating good discriminating

value. The optimal *screening cut-off* was 29, with a sensitivity of 0.79 (95% CI 0.65-0.92) indicating that 79% of those with PTSD will be detected. The specificity was 0.80 (95% CI 0.77-0.82), indicating that there is an 80% probability that those who do not have PTSD will score below the cut-off. This cut-off resulted in 9897 false positive and 359 false negative diagnoses. In contrast, the optimal *epidemiological cut-off* was 53, with a specificity of 0.97 (95% CI 0.97-0.98), but a sensitivity of 0.25 (95% CI 0.15-0.35). This cut-off resulted in 1215 false positive and 1247 false negative diagnoses. Using its established cut-off of 50, the PCL-C showed very low sensitivity (0.30, 95% CI 0.19-0.40) though high specificity (0.97 95% CI 0.96-0.97).

K10

Table 6 presents optimal K10 scores for detecting 30-day ICD-10 anxiety/affective disorder. The AUROC for detecting *any 30-day anxiety disorder* was 0.75 (95% CI 0.60-0.89), indicating fair to good discriminating value. The optimal *screening cut-off* was 17: sensitivity was 0.68 (95% CI 0.49-0.87), indicating this cut-off would detect 68% of those with anxiety disorder, and specificity was 0.72 (95% CI 0.68-0.75), indicating that 72% of those without anxiety disorder will score below this cut-off. This cut-off resulted in 13115 false positive and 1210 false negative diagnoses. In contrast, the optimal *epidemiological cut-off* was 26, with a specificity of 0.95 (95% CI 0.93-0.96) (but a sensitivity of 0.30 (95% CI 0.19-0.40)). This cut-off resulted in 2470 false positive and 2674 false negative diagnoses.

The AUROC for detecting *any 30-day ICD-10 affective disorder* was 0.81 (95% CI 0.70-0.91), indicating good discriminating value. The optimal *screening cut-off* was 19: sensitivity was 0.75 (95% CI 0.59-0.91), and specificity was 0.79 (95% CI 0.76-0.82). This cut-off resulted in 10207 false positive and 336 false negative diagnoses. In contrast, the optimal *epidemiological cut-off* was 31 (specificity = 0.98 (95% CI 0.97-0.98), sensitivity = 0.23 (95% CI 0.13-0.33)). This cut-off resulted in 1117 false positive and 1021 false negative diagnoses. Both of these cut-offs were slightly higher than the cut-off scores identified for any anxiety disorder.

Finally, the AUROC for detecting *any 30-day ICD-10 anxiety or affective disorder* was 0.75 (95% CI 0.63-0.86), indicating fair to good discriminating value (see Figure 1). The optimal screening cut-off was 19: sensitivity was 0.59 (95% CI 0.44-0.73), and specificity was 0.81 (95% CI 0.78-0.84). This cut-off resulted in 8530 false positive and 1883 false negative diagnoses. In contrast, the optimal *epidemiological cut-off* was 25, with a specificity of 0.93 (95% CI 0.92-0.95) (but a sensitivity of 0. 30 (95% CI 0.21-0.39)). This cut-off resulted in 2974 false positive and 3169 false negative diagnoses.

Overall, conservative screening and epidemiological cut-off values of 17 and 25 (respectively) would identify both anxiety and affective disorders within the ADF population.

Discussion

This is the first study to test the diagnostic validity of three routinely-used mental health screening scales in a large representative military sample. All scales showed good to excellent levels of overall diagnostic validity, and more specifically, their optimal screening cut-offs could sensitively detect disorder whilst maintaining good specificity (although the degree to which each scale did this differed). In most cases, these screening cut-offs paralleled those already established in other populations. In sum, these scales appear to be useful for military personnel.

The AUDIT showed excellent discriminating ability between military personnel with and without alcohol disorders, particularly alcohol dependence. This is consistent with research in various populations, including Australian Navy personnel (Degenhardt et al., 2001; McKenzie et al., 2006; Reinert & Allen, 2002). The optimal screening cut-off of 8 is identical to that recommended by the World Health Organization, used in military research (Fear et al., 2010), and used for ADF screening, though it was slightly lower than the optimal cut-off found in Australian Navy personnel (McKenzie et al., 2006). This cut-off showed excellent sensitivity and good specificity: thus, while the AUDIT may detect the majority of personnel with alcohol disorders (having only missed 34 ADF members within our analyses), it will require rigorous follow-up interview procedures to screen out the significant number of false

positives (up to 13000 members). Our results support the AUDIT's use (and the cut-off of 8) in military personnel.

The PCL-C also showed good discrimination between personnel with and without PTSD, consistent with previous research (McDonald & Calhoun, 2010). The optimal screening cut-off of 29 showed a balance of good sensitivity and specificity; results were similar to those in actively-serving personnel and primary care veterans (Bliese et al., 2008; McDonald & Calhoun, 2010). This cut-off was also similar to the current ADF cut-off of 30 (Department of Defence, 2009; Nicholson, 2006). However, the originally-recommended cut-off of 50 (Weathers et al., 1993) appeared too high, and did not identify most personnel with PTSD. Bliese et al. (2008) speculated that perceived stigma might explain lower optimal cut-offs in primary care samples (like theirs) and post-deployment settings, compared with treatment-seeking or anonymous epidemiological samples. However, this reasoning cannot be applied to our study, as participation was voluntary and results confidential. Perhaps because many personnel knew the questionnaire from post-deployment screening, they completed it in the same manner, as if it was not confidential. Regardless, it appears that the current ADF screening cut-off of 30 performs well, although a slight score reduction may improve the proportion of correct diagnoses.

The K10 was the least effective, despite showing reasonable diagnostic accuracy overall, with good ability for predicting affective disorders and lower though fair ability for predicting anxiety disorders. This relatively poorer performance was perhaps unsurprising given the K10 was designed to measure non-specific distress rather than any particular disorder (Kessler et al., 2002). However, the K10 has performed excellently in predicting anxiety and mood disorders in community populations (Andrews & Slade, 2001; Furukawa et al., 2003; Kessler et al., 2002; Kessler et al., 2003; Oakley Browne et al., 2010). Perhaps our sample was more likely to suffer from non-specific pathology not reflected in formal diagnoses. In the only study assessing cut-offs, the range of 'optimal screening cut-offs' (using our criterion) was 16-18, spanning our cut-off of 17 (Andrews & Slade, 2001); however, our cut-off demonstrated lower sensitivity. Our optimal cut-off is also similar to the cut-off of 20 used by

Australian primary care clinics and the ADF. Although this score difference was only slight, the established cut-off of 20 showed particularly lower sensitivity, only detecting about half of ADF members with anxiety/affective disorders. Thus, the established cut-off could be slightly lowered to improve the proportion of correct classifications.

Compared with the optimal screening cut-offs, the diagnostic properties of the optimal epidemiological cut-offs highlight the different purposes of these cut-offs. While the epidemiological cut-offs minimised the number of incorrect diagnoses, as our sample showed low disorder prevalence they favoured specificity, and the resultant poor sensitivity illustrated these cut-offs are not suited for screening as they are poor predictors of disorders for individuals. As case in point, the optimal PCL-C epidemiological cut-off of 53 was very close to the originally-recommended screening cut-off of 50, suggesting that the best use for the original cut-off may be for estimating true population disorder prevalence within military epidemiological research, rather than for screening purposes (as also concluded by Bliese et al., 2008). However, given the low population PTSD prevalence (3.4%), slightly increasing the epidemiological cut-off to 53 would be recommended in order to avoid overestimating prevalence in military personnel (see also McDonald & Calhoun, 2010; Terhakopian et al., 2008). The epidemiological cut-off could also be used to develop screening score bandings to triage personnel, and more efficiently target resources. That is, personnel scoring between screening and epidemiological cut-offs may need cautious and detailed follow-up to determine the presence of disorder. Many of these personnel may experience transient rather than severe problems (see Wright et al., 2005), especially in the immediate decompression phase, and may be quickly returned to duty, with regular follow-up rather than extensive treatment indicated. However, personnel scoring above epidemiological cut-offs might be more quickly referred to appropriate services.

Clinical Implications

These results have important clinical implications for military mental health screening. While we defined optimal screening cut-offs as those that maximised the proportion of correct diagnoses (i.e.,

optimal statistically), optimal cut-offs must ultimately be decided according to each user's needs. Thus, rather than being prescriptive, our results may guide clinical decision-making regarding the best use of these scales. For example, although the optimal K10 cut-off resulted in the highest proportion of total correct classifications, it favoured specificity, showing less than ideal sensitivity. Given that sensitivity is generally preferred for screening purposes, as false positives may be identified through follow-up interviews, a lower cut-off score may be preferred for military personnel. A preference for early intervention and managing potential under-reporting might also lead military forces to consider lower cut-offs. However, in the ADF, a K10 cut-off with a sensitivity of .80 or higher would come at the cost of 16000 additional false positives (assuming an ADF population of 50049, and that all ADF personnel were screened). Thus, lowering cut-offs would necessitate rigorous follow-up procedures to detect these false positives while maintaining high confidentiality to avoid exposure to social stigma, and would require greater follow-up resources. Alternatively, the potential experience of stigma among disorder-free personnel may be considered too great a cost to increase sensitivity at the expense of specificity, especially if an increased likelihood of screening positive reduced personnel 'buy-in'. These and other potential costs and benefits must be considered when selecting cut-offs for military personnel.

Though beyond the scope of our study, scales must also be considered in relation to external criteria to determine their effectiveness within clinical military screening contexts. Importantly, screening will only be optimally effective if it is considered acceptable by personnel. Although ADF post-deployment screening is, with some exceptions, compulsory regardless of acceptability, personnel may be less inclined to report honestly, and to engage in suggested treatment options if they have low faith in the process. Thus, broader cultural issues like confidentiality, stigma, career repercussions, and availability and efficacy of treatment options should also be assessed for potential improvement. It is possible that by improving personnel's trust in the management of mental disorders, such positive cultural shifts may even slightly reduce reliance on screening, if personnel who recognise they have problems feel free to seek help.

Limitations

Several limitations must be considered. First, participants' results were not released to the ADF. While confidentiality is essential for maintaining ethical standards, this study does not parallel the circumstances of ADF screening, where members' results are disclosed to medical officers, and thus can have career implications. It is possible that slightly lower cut-offs may have been found had we conducted this study within standard military screening, as members may under-report symptoms if they believe that negative consequences will result from screening positive.

Moreover, as our procedures differed from those in the ADF, diagnostic validity may differ somewhat. In our study, interviewers with undergraduate degrees (i.e., not clinically trained) and blind to participants' screening scores administered structured diagnostic interviews. In contrast, for post-deployment screening, ADF psychologists and psychological assistants (ADF-trained, not clinically trained) conduct semi-structured interviews based on personnel's screening responses, and follow basic guidelines and personal judgement when making decisions. While some aspects of our study may over-estimate real-life validity, others may underestimate it, and determining the net effect is difficult. Future studies could replicate the actual military experience, and assess the scales' ability to predict referral by ADF psychologists in the context of full disclosure. It is possible that scale validity may be lower in real-life screening contexts (Rona et al., 2004).

As the demographic characteristics of the ADF population were known, our use of inverse probability weighting means that results are representative of the entire ADF population. Of course, the weighting process can contain a degree of error as it involves statistically estimating population-level data from available responders. However, our two-phase design and stratification strategy reduces the possibility of error and improves prevalence estimate precision by focussing diagnostic interviews on the respondents most likely to have a disorder. Additionally, because the interviewees were drawn from the large proportion of the ADF population who completed the Phase 1 questionnaire, the potential for sampling error was further reduced.

These population-based results have important implications for military forces internationally. While other forces show different contextual features from the ADF (e.g., US forces have on average younger personnel and fewer officers, and have experienced longer deployments: Fear et al., 2010; Sundin et al., 2013) we would suggest that generalisability of our results would not be substantially impacted given (1) our disorder prevalence rates are not dissimilar to those from other whole-of-population military studies (e.g., Sareen et al., 2007; Riddle et al., 2007; Fear et al., 2010) when accounting for the differing assessment methodologies used, and (2) our PCL results resemble those documented by Bliese and colleagues (2008) in US soldiers. While it is common practice for researchers and clinicians to use cut-offs that were derived in different countries (see Sundin et al., 2013; Terhakopian et al., 2008), we and others recommend that cut-offs always be validated before use in the population of interest to ensure that the scales are working optimally (McDonald & Calhoun, 2010; Terhakopian et al., 2008).

In conclusion, these three scales represent promising options for military screening. However, only long-term follow-up within military contexts can determine if their use results in referral to and uptake of needed services, and the reduction of mental disorders.

Acknowledgements

We thank Dr. Alan Verhagen for his contribution to the design of this project, Ms. Michelle Lorimer for her statistical assistance, and the many research officers and interviewers who assisted with data collection. This research was funded by the Department of Defence.

Competing Interests

HB, NS and SEH received salaries from the Department of Defence in the last 5 years. Views and opinions expressed within this report are those of the authors, and do not necessarily reflect those of the Department of Defence. While the Department Defence was involved with the study design and data collection, it had no role in the analysis and interpretation of the data, or the decision to submit the manuscript for publication.

Authors' Contributions

AKS conducted literature reviews and drafted the manuscript. MVH designed the study, obtained the data, interpreted the data, and critically reviewed the manuscript. ACM designed the study, interpreted the data, and critically reviewed the manuscript. AKFS assisted in literature review and interpretation of the data, and critically reviewed the manuscript. CED designed the study, analysed and critically interpreted the data, and critically reviewed the manuscript. SEH designed the study, interpreted the data, and critically reviewed the manuscript. HB designed the study, and critically reviewed the manuscript. NS designed the study, and critically reviewed the manuscript.

References

Allen, J.P., Litten, R.Z., Fertig, J.B., Babor, T. (1997). A review of research on the Alcohol Use Disorders Identification Checklist (AUDIT). *Alcohol Clin Exp Res, 21*, 613-619.

Andrews, G., Slade, T. (2001). Interpreting scores on the Kessler Psychological Distress Scale (K10). *Aust NZ J Public Health, 25*, 494-497.

Australian Bureau of Statistics. (2003). Information paper: Use of the Kessler Psychological Distress Scale in ABS health surveys, Australia 2001. Retrieved from http://www.abs.gov.au/ausstats/abs@.nsf/mf/4817.0.55.001.

Babor, T., Higgins-Biddle, J.C., Saunders, J., Monteiro, M.G. (2001). AUDIT - The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Care (2nd ed.). World Health Organization: Department of Mental Health and Substance Dependence, Geneva.

Baillie, A.J. (2005). Predictive gender and education bias in Kessler's psychological distress scale (K10). *Soc Psychiatry Psychiatr Epidemiol, 40*, 743-748.

Barrett, D.H., Doebbeling, C.C., Schwartz, D.A., Voelker, M.D., Falter, K.H., Woolson, R.F., Doebbeling, B.N. (2002). Posttraumatic stress disorder and self-reported physical health status among U.S. military personnel serving during the Gulf War period: A population-based study. *Psychosomatics, 43*, 195-205.

Bleier, J., McFarlane, A.C., McGuire, A., Treloar, S., Waller, M., Dobson, A. (2011). Risk of adverse health outcomes associated with frequency and duration of deployment with the Australian Defence Force. *Mil Med, 176*, 139-146.

Bliese, P.D., Wright, K.M., Adler, A.B., Cabrera, O., Castro, C.A., Hoge, C.W. (2008). Validating the Primary Care Posttraumatic Stress Disorder Screen and the Posttraumatic Stress Disorder Checklist with soldiers returning from combat. *J Consult Clin Psychol, 76*, 272-281.

Degenhardt, L., Conigrave, K., Wutzke, S., Saunders, J. (2001). The validity of an Australian modification of the AUDIT questionnaire. *Drug Alcohol Rev, 20*, 143-154.

Department of Defence. (2008). Defence Instruction (General) Personnel 16–28. Operational mental

health screening. Department of Defence, Canberra.

Department of Defence. (2009). Health Bulletin No 11/2009. Mental health screen for casework. Department of Defence, Canberra.

Dunt, D. (2009). Review of mental health care in the ADF and transition through discharge. Retrieved from http://www.defence.gov.au/health/DMH/Review.htm.

Erbes, C.R., Meis, L.A., Polusny, M.A., Compton, J.S. (2011). Couple adjustment and posttraumatic stress disorder symptoms in National Guard veterans of the Iraq war. *J Fam Psychol, 25*, 479-487.

Fear, N.T., Jones, M., Murphy, D., Hull, L., Iversen, A.C., Coker, B., Machell, L., Sundin, J., Woodhead, C., Jones, N., Greenberg, N., Landau, S., Dandeker, C., Rona, R.J., Hotopf, M., Wessely, S. (2010). What are the consequences of deployment to Iraq and Afghanistan on the mental health of the UK armed forces? A cohort study. *Lancet, 375*, 1783-1797.

French, C., Rona, R., Jones, D., Wessely, S. (2004). Screening for physical and psychological illness in the British Armed Forces: II: Barriers to screening – learning from the opinions of Service personnel. *J Med Screen, 11*, 153-157.

Furukawa, T.A., Kessler, R.C., Slade, T., Andrews, G. (2003). The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychol Med, 33*, 357-362.

Gould, M., Adler, A., Zamorski, M., Castro, C., Hanily, N., Steele, N., Kearney, S., Greenberg, N. (2010). Do stigma and other perceived barriers to mental health care differ across Armed Forces? *J R Soc Med*, 103, 148-156.

Haro, J.M., Saena, A.-B., Brugha, T.S., De Girolamo, D., Guyer, M.E., Jin, R., Lepine, J.P., Mazzi, F., Reneses, B., Vilagut, G., Sampson, N.A., Kessler, R.C. (2006). Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health Surveys. *Int J Methods Psychiatr Res, 15*, 167-180.

Hides, L., Lubman, D.I., Devlin, H., Cotton, S., Aitken, C., Gibbie, T., Hellard, M. (2007). Reliability and validity of the Kessler 10 and Patient Health Questionnaire among injecting drug users. *Aust NZ J*

Psychiatr, 41, 166-168.

Hoge, C.W., Auchterlonie, J.L., Milliken, C.S. (2006). Mental health problems, use of mental health services, and attrition from military service after returning from deployment to Iraq or Afghanistan. *JAMA 295*, 1023-1032.

Hoge, C.W., Castro, C.A., Messer, S.C., McGurk, D., Cotting, D.I., Koffman, R.L. (2004). Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *N Engl J Med, 351*, 13-22.

Hoge, C.W., Lesikar, S.E., Guevara, R., Lange, J., Brundage, J.F., Engel, C.C., Messer, S.C., Orman, D.T. (2002). Mental disorders among U.S. military personnel in the 1990s: Association with high levels of health care utilization and early military attrition. *Am J Psychiatr, 159*, 1576-1583.

Hoge, C.W., Toboni, H.E., Messer, S.C., Bell, N., Amoroso, P., Orman, D.T. (2005). The occupational burden of mental disorders in the U.S. military: Psychiatric hospitalizations, involuntary separations, and disability. *Am J Psychiatr 162*, 585-591.

Iversen, A., van Staden, L., Hughes, J.H., Greenberg, N., Hotopf, M., Rona, R.J., Thomicroft, G., Wessely, S., Fear, N.T. (2011). The stigma of mental health problems and other barriers to care in the UK Armed Forces. *BMC Health Services Res*, *11*, 31-40.

Iversen, A.C., Fear, N.T., Ehlers, A., Hughes, J.H., Hull, L., Earnshaw, M., Greenberg, N., Rona, R., Wessely, S., Hotopf, M. (2008). Risk factors for post-traumatic stress disorder among UK Armed Forces personnel. *Psychol Med, 38*, 511-522.

Kessler, R.C., Andrews, G., Colpe, L.J., Hiripi, E., Mroczek, D.K., Normand, S.L.T., Walters, E.E., Zaslavsky, A.M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol Med, 32*, 959-976.

Kessler, R.C., Barker, P.R., Colpe, L.J., Epstein, J.F., Gfroerer, J.C., Hiripi, E., Howes, M.J., Normand, S.L.T., Manderscheid, R.W., Walters, E.E., Zaslavsky, A.M. (2003). Screening for serious mental illness in the general population. *Arch Gen Psychiatr, 60*, 184-189.

Kessler, R.C., Üstün, T.B. (2004). The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int J Methods*

Psychiatr Res, 13, 93-121.

Kim, P. Y., Thomas, J. L., Wilk, J. E., Castro, C. A., & Hoge, C. W. (2010). Stigma, barriers to care, and use of mental health services among active duty and Naitonal Guard soliders after ocmbat. *Psychaitric Services, 61*, 572-588.

McDonald, S.D., Calhoun, P.S. (2010). The diagnostic accuracy of the PTSD checklist: A review. *Clin Psychol Rev, 30*, 976-987.

McFarlane, A.C., Hodson, S.E., Van Hooff, M., Davies, C. (2011). Mental health in the Australian Defence Force: 2010 ADF Mental Health and Wellbeing Study: Full report. Department of Defence, Canberra.

McKenzie, D.P., McFarlane, A.C., Creamer, M., Ikin, J.F., Forbes, A.B., Kelsall, H.L., Clarke, D.M., Glass, D.C., Ittak, P., Sim, M.R. (2006). Hazardous or harmful alcohol use in Royal Australian Navy veterans of the 1991 Gulf War: Identification of high risk subgroups. *Addict Behav, 31*, 1683-1694.

Nicholson, C. (2006). A review of the PTSD Checklist. Research Report 08/2006. Department of Defence, Canberra.

Oakley Browne, M.A., Wells, J.E., Scott, K.M., McGee, M.A., for the New Zealand Mental Health Survey Research Team. (2010). The Kessler Psychological Distress Scale in Te Rau Hinengaro: The New Zealand Mental Health Survey. *Aust NZ J Psychiatr, 44*, 314-322.

Reinert, D.F., Allen, J.P. (2002). The Alcohol Disorders Use Identification Test (AUDIT): A review of recent research. *Alcohol Clin Exp Res, 26*, 272-279.

Riddle, J.R., Smith, T.C., Smith, B., Corbeil, T.E., Engel, C.C., Wells, T.S., Hoge, C.W., Adkins, J., Zamorski, M., Blazer, D., for the Millenium Cohort Study Team. (2007). Millennium Cohort: The 2001 - 2003 baseline prevalence of mental disorders in the U.S. military. *J Clin Epidemiol, 60*, 192-201.

Rona, R., Hooper, R., Jones, D., French, C., Wessely, S. (2004). 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. *J Med Screen, 11*, 158-161.

Rona, R., Hyams, K.C., Wessely, S. (2005). Screening for psychological illness in military personnel.

JAMA 293, 1257-1260.

Rona, R., Jones, M., Iversen, A., Hull, L., Greenberg, N., Fear, N.T., Hotopf, M., Wessely, S. (2009). The impact of posttraumatic stress disorder on impairment in the UK military at the time of the Iraq war. *J Psychiatr Res, 43*, 649-655.

Rona, R.J., Jones, M., French, C., Hooper, R., Wessely, S. (2004). Screening for physical and psychological illness in the British Armed Forces: I: The acceptability of the programme. *J Med Screen, 11,* 148-153.

Ryan, M.A.K., Smith, T.C., Smith, B., Amoroso, P., Boyko, E.J., Gray, G.C., Gackstetter, G.D., Riddle, J.R., Wells, T.S., Gumbs, G., Corbeil, T.E., Hooper, T.I. (2007). Millennium Cohort: enrollment begins a 21-year contribution to understanding the impact of military service. *J Clin Epidemiol, 60,* 181-191.

Sareen, J., Cox, B.J., Afifi, T.O., Stein, M.B., Belik, S.-L., Meadows, G., Asmundson, G.J.G. (2007). Combat and peacekeeping operations in relation to prevalence of mental disorders and perceived need for mental health care: Findings from a large representative sample of military personnel. *Arch Gen Psychiatr, 64,* 843-852.

Steele, N., Goodman, M. (2006). History of ADF Mental Health Screening 1999–2005. Psychology Research and Technology Group Technical Brief 4/2006. Department of Defence, Canberra.

Steele, N., Twomey, A., 2008. Post-deployment mental health screning, surveillance and research in TTCP countries. Report for the Technical Cooperation Program Human Resources and Performance group.

Sundin, J., Herrell, R. K., Hoge, C. W., Fear, N. T., Adler, A., Greenberg, N., Riviere, L., Thomas, J., Wessely, A. & Bliese, P. (2013). Mental health outcomes in US and UK military personnel returnin from Iraq. *The British Journal of Psychiatry, 203*, 1-8.

Terhakopian, A., Sinaii, N., Engel, C. C., Schnurr, P. P., & Hoge, C. W. (2008). Estimating population prevalence of posttraumatic stress disorder: An example using the PTSD checklist. *Journal of Traumatic Stress, 21(3*), 290−300.

Thomas, J.L., Wilk, J.E., Riviere, L.A., McGurk, D., Castro, C.A., Hoge, C.W. (2010). Prevalence of

mental health problems and functional impairment among active component and national guard soldiers 3 and 12 months following combat in Iraq. *Arch Gen Psychiatr 67*, 614-623.

Warner, C.H., Appenzeller, G.N., Grieger, T., Belenkiy, S., Breitbach, J., Parker, J., Warner, C.M., Hoge, C. (2011). Importance of anonymity to encourage honest reporting in mental health screening after combat deployment. *Arch Gen Psychiatr, 68*, 1065-1071.

Weathers, F.W., Litz, B.T., Herman, D.S., Huska, J.A., Keane, T.M. (1993). The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility. Annual meeting of the International Society for Traumatic Stress Studies, San Antonio, Texas.

Wilkins, K.C., Lang, A.J., Norman, S.B. (2011). Synthesis of the psychometric properties of the PTSD checklist (PCL) military, civilian, and specific versions. *Depress Anxiety, 28*, 596-606.

World Health Organization. (1992). The International Classification of Diseases: Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines. 10. World Health Organization, Geneva.

Wright, K.M., Bliese, P.D., Adler, A.B., Hoge, C., Castro, C.A., Thomas, J.L. (2005). Letter to the editor: Screening for psychological illness in the military. *JAMA 294*, 41-42.

Wright, K.M., Bliese, P.D., Thomas, J.L., Adler, A.B., Eckford, R.D., Hoge, C. (2007). Contrasting approaches to psychological screening with U.S. combat soldiers. *J Trauma Stress 20*, 965-975.

Wright, K.M., Thomas, J.L., Adler, A.B., Ness, J.W., Hoge, C.W., Castro, C.A. (2005). Psychological screening procedures for deploying US forces. *Mil Med, 170*, 555-562.

Table 1

Demographic characteristics of study respondents and the total ADF population

Variable	Phase 2 study sample (*n* = 1798)	Total ADF population (*n* = 50 049)
Age	38.3 (9.4)	33.2 (9.2)
Male %	75.6	86.4
Service		
Army %	39.8	50.7
Navy %	21.4	23.2
Air Force %	38.8	26.1
Rank		
Commissioned officer %	36.4	24.0
Non-commissioned officer %	49.4	44.6
Other ranks %	14.1	31.4
Time in ADF (years)	16.2 (9.8)	11.6 (8.8)
Been deployed %	61.8	65.4
Married %	77.2	62.9
Highest educational qualifications %		
High school or less	13.1	-[a]
Certificate/diploma	36.9	-[a]
University degree	50.0	-[a]
MEC status		
MEC 1 %	50.4	65.6
MEC 2 %	34.0	23.4
MEC 3 %	12.5	8.9
MEC 4 %	3.2	2.1

MEC = medical employment classification (MEC 1/2 = fit to deploy, MEC 3/4 = unfit to deploy). [a]Unable to determine education of non-responders.

Table 2

Mental health screening scale scores (n = 50 049)

Variable	*M* or % (95% CI)
Levels of mental health problems	
AUDIT	6.0 (5.9, 6.0)
Zone II or above (8+) %	26.4% (25.5, 27.2)
Zone IV or above (20+) %	1.4% (1.2, 1.5)
PCL	22.7 (22.6, 22.8)
At least moderate (30+) %	15.4% (14.7, 16.0)
Very high (50+) %	3.0% (2.8, 3.1)
K10	15.4 (15.3, 15.5)
At least 'moderate' (16+) %	35.4% (34.3, 36.3)
At least 'high' (22+) %	12.9% (12.3, 13.4)

Table 3

Prevalence of CIDI ICD-10 30-day disorder (n = 50 049)

30-day disorder variable	% (95% CI)
Any anxiety *or* affective disorder	9.1% (6.8, 11.4)
Any anxiety disorder	7.5% (5.4, 9.7)
Any affective disorder	2.6% (1.9, 3.4)
Post-traumatic stress disorder	3.4% (2.6, 4.2)
Any alcohol disorder	1.0% (0.5, 1.4)
Alcohol harmful use	0.2% (0.0, 0.5)
Alcohol dependence	0.8% (0.3, 1.2)

Table 4

Properties of the AUDIT for predicting 30-day ICD-10 alcohol disorders

Cut-off	Sensitivity		Specificity		PPV		NPV		Overall efficiency	
	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI
Alcohol harmful use										
7	1.00	1.00–1.00	0.69	0.66–0.72	0.01	0.00–0.01	1.00	1.00–1.00	0.69	0.66 – 0.72
8†	1.00	1.00–1.00	0.75	0.73–0.78	0.01	0.00–0.02	1.00	1.00–1.00	0.75	0.73 – 0.78
9	0.57	0.07–1.00	0.82	0.80–0.84	0.01	0.00–0.01	1.00	1.00–1.00	0.82	0.80 – 0.84
…										
25	0.00	0.00–0.00	1.00	0.99–1.00	0.00	0.00–0.00	1.00	1.00–1.00	0.99	0.99 – 1.00
26§	0.00	0.00–0.00	1.00	1.00–1.00	0.00	0.00–0.00	1.00	1.00–1.00	0.99	0.99 – 1.00
27	0.00	0.00–0.00	1.00	1.00–1.00	0.00	0.00–0.00	1.00	1.00–1.00	0.99	0.99 – 1.00
Alcohol dependence										
8	0.94	0.85–1.00	0.76	0.73–0.78	0.03	0.01–0.04	1.00	1.00–1.00	0.76	0.73–0.78
9†	0.91	0.81–1.00	0.83	0.81–0.85	0.04	0.02–0.06	1.00	1.00–1.00	0.83	0.81–0.85
10	0.80	0.59–1.00	0.86	0.84–0.88	0.04	0.02–0.07	1.00	1.00–1.00	0.86	0.84–0.87
…										
20	0.23	0.01–0.45	0.99	0.99–1.00	0.20	0.01–0.39	0.99	0.99–1.00	0.99	0.98–0.99
21§	0.08	0.01–0.18	0.99	0.99–1.00	0.09	0.02–0.19	0.99	0.99–1.00	0.99	0.98–0.99
22	0.08	0.01–0.18	0.99	0.99–1.00	0.11	0.02–0.23	0.99	0.99–1.00	0.99	0.98–0.99
Any alcohol disorder										
7	1.00	1.00–1.00	0.70	0.66–0.73	0.03	0.02–0.05	1.00	1.00–1.00	0.70	0.67-0.73
8†	0.95	0.89–1.00	0.76	0.73–0.78	0.04	0.02–0.06	1.00	1.00–1.00	0.76	0.73-0.79
9	0.83	0.64–1.00	0.83	0.81–0.85	0.05	0.02–0.07	1.00	1.00–1.00	0.83	0.81-0.85
…										
19	0.21	0.03–0.39	0.99	0.99–0.99	0.17	0.03–0.32	0.99	0.99–1.00	0.98	0.98-0.99
20§	0.19	0.02–0.37	0.99	0.99–1.00	0.22	0.03–0.41	0.99	0.99–1.00	0.99	0.98-0.99
21	0.08	0.00–0.16	0.99	0.99–1.00	0.12	0.00–0.23	0.99	0.99–1.00	0.98	0.98-0.99

Note. †Optimal screening cut-off. § Optimal epidemiological cut-off. PPV = positive predictive value; NPV = negative predictive value. The scores above and below the optimal screening and epidemiological cut-offs are also displayed.

Table 5

Properties of the PCL-C for predicting 30-day ICD-10 post-traumatic stress disorder

Cut-off	Sensitivity		Specificity		PPV		NPV		Overall efficiency	
	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI
28	0.79	0.65–0.92	0.78	0.75–0.80	0.11	0.08–0.13	0.99	0.98–1.00	0.78	0.75–0.80
29†	0.79	0.65–0.92	0.80	0.77–0.82	0.12	0.09–0.15	0.99	0.98–1.00	0.80	0.77–0.82
30	0.74	0.60–0.87	0.82	0.80–0.84	0.12	0.09–0.15	0.99	0.98–1.00	0.82	0.80–0.83
…										
50	0.30	0.19–0.40	0.97	0.96–0.97	0.23	0.15–0.31	0.98	0.97–0.98	0.94	0.93–0.95
…										
52	0.28	0.18–0.39	0.97	0.96–0.98	0.26	0.17–0.35	0.98	0.97–0.98	0.95	0.94–0.96
53§	0.25	0.15–0.35	0.97	0.97–0.98	0.26	0.16–0.36	0.97	0.97–0.98	0.95	0.94–0.96
54	0.21	0.12–0.30	0.98	0.97–0.98	0.24	0.14–0.34	0.97	0.96–0.98	0.95	0.94–0.96

Note. †Optimal screening cut-off. § Optimal epidemiological cut-off. PPV = positive predictive value; NPV = negative predictive value. The scores above and below the optimal screening and epidemiological cut-offs are also displayed (as well as the established cut-off of 50).

Table 6

Properties of the K10 for predicting 30-day ICD-10 anxiety and affective disorders

Cut-off	Sensitivity		Specificity		PPV		NPV		Overall efficiency	
	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI
Any anxiety disorder										
16	0.73	0.53–0.93	0.64	0.60–0.68	0.14	0.12–0.17	0.97	0.93–1.00	0.65	0.61–0.69
17†	0.68	0.49–0.87	0.72	0.68–0.75	0.16	0.13–0.20	0.96	0.94–0.99	0.71	0.68–0.75
18	0.61	0.44–0.79	0.76	0.72–0.79	0.17	0.14–0.21	0.96	0.93–0.99	0.75	0.71–0.78
…										
20	0.49	0.35–0.64	0.84	0.82–0.87	0.21	0.17–0.25	0.95	0.93–0.98	0.82	0.79–0.85
…										
25	0.31	0.21–0.42	0.93	0.92–0.95	0.27	0.20–0.34	0.94	0.92–0.97	0.88	0.86–0.91
26§	0.30	0.19–0.40	0.95	0.93–0.96	0.31	0.23–0.39	0.94	0.92–0.97	0.90	0.87–0.92
27	0.25	0.16–0.34	0.95	0.94–0.96	0.30	0.21–0.38	0.94	0.92–0.96	0.90	0.87–0.92
Any affective disorder										
18	0.76	0.60–0.92	0.74	0.71–0.77	0.07	0.05–0.09	0.99	0.98–1.00	0.74	0.71–0.77
19†	0.75	0.59–0.91	0.79	0.76–0.82	0.09	0.06–0.11	0.99	0.98–1.00	0.79	0.76–0.82
20	0.69	0.54–0.85	0.83	0.81–0.85	0.10	0.07–0.13	0.99	0.98–1.00	0.83	0.81–0.85
…										
30	0.28	0.17–0.40	0.97	0.97–0.98	0.22	0.13–0.31	0.98	0.97–0.99	0.95	0.94–0.96
31§	0.23	0.13–0.33	0.98	0.97–0.98	0.21	0.13–0.30	0.98	0.97–0.99	0.95	0.95–0.97
32	0.17	0.10–0.25	0.98	0.98–0.99	0.20	0.12–0.28	0.98	0.97–0.99	0.96	0.95–0.97
Any anxiety *or* affective disorder										
18	0.62	0.46–0.77	0.76	0.73–0.80	0.21	0.17–0.24	0.95	0.92–0.98	0.75	0.71–0.78
19†	0.59	0.44–0.73	0.81	0.78–0.84	0.24	0.19–0.28	0.95	0.92–0.98	0.79	0.76–0.82
20	0.50	0.37–0.63	0.85	0.83–0.87	0.25	0.21–0.29	0.94	0.92–0.97	0.82	0.79–0.85
…										
24	0.35	0.25–0.45	0.93	0.91–0.94	0.33	0.26–0.40	0.93	0.91–0.96	0.88	0.85–0.90
25§	0.30	0.21–0.39	0.93	0.92–0.95	0.32	0.24–0.39	0.93	0.91–0.96	0.88	0.85–0.90
26	0.28	0.20–0.37	0.95	0.94–0.96	0.36	0.27–0.44	0.93	0.91–0.95	0.89	0.86–0.91

Note. †Optimal screening cut-off. § Optimal epidemiological cut-off. PPV = positive predictive value; NPV = negative predictive value. The scores above and below the optimal screening and epidemiological cut-offs are also displayed (as well as the current ADF cut-off of 20).