

***Identification of key depressive symptoms and their
comorbidity with physical and psychological illness in
Australian Gulf War Veterans***

Technical Report

Department of Veterans' Affairs

Applied Research Program 1127

Research Team

Dr Dean McKenzie, Principal Investigator^{1,2}

Dr Helen Kelsall¹

Professor David Clarke³

Professor Andrew Forbes¹

Dr Jillian Ikin¹

Professor Malcolm Sim¹

¹Department of Epidemiology and Preventive Medicine, Monash University

²Centre for Adolescent Health, Murdoch Childrens Research Institute, Royal Children's Hospital

³Discipline of Psychiatry, Monash University

Table of Contents

1	List of Abbreviations	2
2	Acknowledgements	4
3	Introduction	5
3.1	Depression in veterans	5
3.2	Depression and physical health	5
3.3	Identification of depression and screening tests	6
4	Aims of study	8
5	Study methodology	9
5.1	Study population	9
5.2	Instruments	9
5.2.1	Depressive symptomatology	9
5.2.2	Depression diagnosis	10
5.2.3	Comorbid physical and psychological illness	11
5.3	Statistical analyses	12
5.3.1	Stepwise logistic regression	12
5.3.2	Decision trees	13
5.3.3	Decision rules	13
5.3.4	Threshold or cut-off score	13
5.3.5	Screening and diagnostic performance of the different methods	14
5.3.6	Comorbid physical and psychological illness	14
5.3.7	Possible confounding factors	15
5.3.8	Statistical software	16
6	Results	17
6.1	Individual and simultaneous relationships between GHQ-12 items and major depression	17
6.2	CART decision tree and SPAN decision rule for key GHQ-12 items	20
6.3	Twelve item cut-off / threshold for GHQ-12 total score	24
6.4	Screening and diagnostic performance of the above methods	24
6.5	Brief depression screen and medical and psychological conditions	26
6.6	Comorbid CART brief depression screen and medical / psychological conditions and health-related quality of life	28
7	Discussion	31
7.1	Key depressive symptoms	31
7.2	Screening and diagnostic performance of brief depression screen	32
7.3	Brief depression screen and physical and psychological conditions	32
7.4	Comorbid brief depression screen and physical / psychological conditions and health-related quality of life	33
7.5	Implications for identification of depression in veterans	33
7.6	Implications for further research	35
7.7	Conclusion	36
8	Appendices	37
8.1	Appendix 1: 12 item General Health Questionnaire (GHQ-12)	37
8.2	Appendix 2: DSM-IV (Diagnostic and Statistical Manual, version 4) criteria for major depression	38
8.3	Appendix 3: 12 item Short-Form Health Survey (SF-12)	39
9	References	41

1 List of Abbreviations

AUC	Area under the Receiver Operating Characteristic (ROC) Curve
BMI	Body mass index, a measure of body fat based on height and weight
CART	Classification and Regression Tree Analysis, a computational technique for identifying subgroups of individuals. In this study the subgroups of individuals are defined by combinations of GHQ-12 symptoms that have similar levels of DSM-IV major depression
CI	Confidence interval, the probability that a population parameter such as an odds ratio, will fall between two specified values
CIDI	Composite International Diagnostic Interview, a structured diagnostic interview for obtaining information necessary for making diagnoses for psychiatric disorders such as major depression, in epidemiological studies
DSM-IV	Diagnostic and Statistical Manual, version 4; a commonly employed set of formal diagnostic criteria for psychiatric disorders such as major depression
DSM-5	Diagnostic and Statistical Manual, version 5, released in 2013
GHQ-12	General Health Questionnaire, 12 item version; a measure of psychological distress
GHQ-28	General Health Questionnaire, 28 item version, a measure of psychological distress.
GHQ-60	General Health Questionnaire, 60 item version, a measure of psychological distress
GP	General Practitioner
K6	Kessler Psychological Distress Scale, six item version
K10	Kessler Psychological Distress Scale, ten item version
MCS-12	Mental Component Summary, mental health component of the SF-12
NPV	Negative predictive value, probability of not meeting diagnostic criteria for those with a negative screening test result
OR	Odds ratio, the ratio of the odds of an event occurring in one group to the odds of it occurring in another group, (e.g., the ratio of the odds of major depression occurring in those with a particular depressive symptom present to the odds of major depression occurring in those not having a particular depressive symptom present)
PCS-12	Physical Component Summary, physical health component of the SF-12
PHQ-9	The 9 item depression scale of the Patient Health Questionnaire

PPV	Positive predictive value, probability of meeting diagnostic criteria for those with a positive screening test result
RAAF	Royal Australian Air Force
RAN	Royal Australian Navy
ROC	Receiver Operating Characteristic, a measure of overall screening and diagnostic performance of a particular screening test or model
SF-12	Short-Form Health Survey, 12 item version; a measure of well-being and health-related quality of life
SPAN	Search Partition Analysis, a computational technique for developing decision rules, (e.g., combinations of GHQ-12 symptoms associated with DSM-IV major depression)

2 Acknowledgements

The present study of the physical and psychological health of Australian Gulf War veterans, Applied Research Program project number 1127, was funded by the Australian Government Department of Veterans' Affairs. We acknowledge the assistance of Dr Eileen Wilson, Mr Tim Cummins, and Ms Sandy Bell.

The study is part of the Australian Gulf War Veterans' Health Study, which was commissioned and funded by the Australian Government Department of Veterans' Affairs.

The research team also acknowledges Ms Jane Miosge and Dr Jed Blore for research assistance and administrative support. We acknowledge the contribution of the participants in the Australian Gulf War Veterans' Health Study.

3 Introduction

3.1 Depression in veterans

Awareness and understanding of war-related psychological health issues are key considerations for defence and veterans' organisations. This is because of the short-term effects of psychological ill-health on combat readiness, organisational productivity and service retention, as well as the longer term requirements for acute and ongoing care and compensation where applicable.

Major depression can be defined as being 'lowered mood [that] persists, brings intense distress, and interferes with the person's ability to cope with the ordinary demands of living, whether of work, study, or personal relationships' (page 83) (1). Major depression is a crucial, but under-recognised, problem in serving military, as well as veteran populations (2, 3).

Affective or mood disorder, including major depression, was found to be the second most common post-war psychological disorder in Australian 1990/91 Gulf War veterans after substance abuse disorders (4). Moreover, affective disorder occurs over twice as frequently as anxiety disorders, including post-traumatic stress disorder (PTSD) (4), in this population. High rates of depression have also been observed in veterans of recent conflicts in Afghanistan and Iraq (5), as well as in veterans of earlier conflicts, such as the Korean War (6).

3.2 Depression and physical health

General Australian population figures suggest that depression has the third highest burden of disease, in terms of years of life lost due to ill-health, disability or early death, and is the primary cause of non-fatal disability (7). The World Health Organization reports that depression is also the third highest burden of disease internationally, and estimates that depression will be the number one global health concern by 2030 (8).

In addition to being a major public health issue in its own right, depression is associated with both physical and psychological illness. For example, physical illnesses, such as coronary heart disease and diabetes, are risk factors for depression in veterans, while depression is itself a risk factor for these diseases (9). Multisymptom illness co-occurs, (occurs comorbidly), with depression in Gulf War veterans (10), as do other psychological disorders such as

posttraumatic stress disorder and alcohol use disorders (11). Finally, depression has been found to be associated with lower health-related quality of life in veterans (12, 13), particularly when accompanied by comorbid psychological disorders such as anxiety (14).

3.3 Identification of depression and screening tests

Depression can be very difficult to recognise in practice. In the general primary health care setting such as General Practitioner (GP) clinics for example, only around half of those persons with major depression are actually identified (15). In order to aid identification of possible cases of depression, self-reported screening or case-identification instruments are often employed. As well as being able to recognise potential depression, such instruments need to be easy to score, comprehensive, clinically relevant, and brief (16). The latter requirement is particularly important when large numbers of people need to be assessed, or when screening needs to take place in busy medical practice settings.

Case identification of depression using brief self-report instruments is generally based upon determining the total number of depressive symptoms that are present, rather than which specific depressive symptoms are most highly associated with a diagnosis of major depression (17).

There has recently been a renewal of interest in attempting to gain a better understanding of depression by considering its actual symptoms (18, 19). For example, veterans have often been found to experience some depressive symptoms such as feelings of worthlessness, but not others, such as loss of interest or pleasure (20).

A small number of veteran studies have focussed on specific depressive symptoms, such as worthlessness, e.g. (20), but have not sought to identify the most important symptoms in terms of their association with a diagnosis of depression.

General population studies (21), including our own (18, 22, 23) have attempted to develop brief screening tests from longer, existing instruments such as the General Health Questionnaire (GHQ), a measure of psychological distress (24). Brief screening instruments for depression in the general population, consisting of four items (22), two items (5), or even one item (25), have been described. Two item tests, often used with veteran populations (5), generally consist of depressed mood and anhedonia (lack of interest or pleasure), and so do

not measure other key depressive symptoms in terms of their association with an actual diagnosis of depression.

Small numbers of key depressive symptoms may be easily recognisable by veterans themselves, as well as by their family members and health providers such as medical practitioners. The presence of such symptoms may therefore act as an early warning system for the possible presence of major depression, without first needing to obtain all the symptom information normally needed to make such a diagnosis.

4 Aims of study

The present study aimed to identify a small number of key symptoms of major depression, which could readily be recognised by veterans, their caregivers and their health providers. We then examined the relationship between the presence of these key symptoms, and physical and psychological illness or disorders, as well as health-related quality of life.

More specifically, the present study aimed to:

- identify a small number of key depressive symptoms which are highly associated with a Diagnostic and Statistical Manual, version 4 (DSM-IV) diagnosis of major depression (26)
- directly compare the performance of a brief depression case screening or identification approach, based upon identifying a small number of key symptoms (i.e., the quality or type of symptoms present), with that based on the common practice of developing a threshold or cut-off score using the total number of items (i.e., the quantity of symptoms present)
- Compare case definition of the key symptom approach and the threshold approach on their identification of major depression, as well as on their association with physical and psychological disorders, and health-related quality of life.

5 Study methodology

5.1 Study population

As part of our ongoing research into the physical and psychological health of Australian Gulf War veterans and military comparison group, this study employed the dataset of the Australian Gulf War Veteran's Health Study, a cross-sectional study of 1456 Australian veterans of the 1990/91 Gulf War (80.5% response rate) and a randomly sampled age, service- and rank-matched military comparison group, undertaken by our research team in 2000-2002. In that study, comprehensive demographic, physical and psychological health information, including depression diagnosis and symptoms, was collected by postal questionnaire, as well as through face to face interviews with medical practitioners and psychologists (4). The present study focused on the veterans, and involved analysis of existing data.

Due to the small number of female Gulf War veterans (n=38), male Australian Army (n=87) and male Royal Australian Air Force (RAAF) (n=105) Gulf War veterans, analyses were confined to the 1204 (97.7%) of the 1232 male Royal Australian Navy (RAN) Gulf War veterans with CIDI diagnostic data. The age of the 1204 male RAN Gulf War veterans at the time of data collection (2000-2003) ranged from 27.6 to 61.5 years, with an average age of 37.3 years, and a standard deviation of 6.1 years.

5.2 Instruments

5.2.1 Depressive symptomatology

Depressive symptoms were assessed using the twelve item version of the General Health Questionnaire (GHQ-12) (24), which has been employed in a wide variety of studies, including those of veterans from different deployments (12, 27). The GHQ-12 has been shown to exhibit comparable screening performance to the longer versions of the GHQ, such as the twenty eight item GHQ-28 and the sixty item GHQ-60 (28).

The GHQ-12 consists of items such as "have you recently been feeling unhappy and depressed", each coded on a four-point scale, the specific wording of which differs across items, as shown in Appendix 1.

The various versions of the GHQ were developed as community research tools, designed to identify general psychiatric illness, appropriate to general medical care, rather than to identify any particular psychiatric disorder such as depression or anxiety. Although widely employed as a self-report measure of general psychological distress (29), the GHQ-12 has also been used as a self-report measure of depression, performing comparably to specific self-report measures of this disorder (30).

In the present study, as well as identifying key specific depression symptoms, or combinations thereof, associated with a diagnosis of major depression, we also identified a cut-off or threshold score for the presence of major depression based upon the total GHQ-12 score. We employed the standard, or binary coding of items (24), in which the first two categories were combined into a response of “absent”, the latter two categories were combined into a response of “present”, and the resulting binary items summed in order to provide a total score.

5.2.2 Depression diagnosis

Diagnosis of major depression was obtained using DSM-IV criteria (26). These criteria, also employed in the recently released DSM-5 (31), require the presence of at least five symptoms, including depressed mood or loss of interest or pleasure, out of a possible nine symptoms shown in Appendix 2 (26). Such symptoms must be present for a period of at least two weeks.

Diagnostic criteria were applied using the computer-assisted Composite International Diagnostic Interview (CIDI) version 2.1 (32), administered by psychologists during the medical assessment of the Gulf War Veterans' Health Study.

The CIDI only collects further information on depressive symptoms such as feelings of worthlessness if the core criteria of depressed mood and/or loss of interest or pleasure are met. These core symptoms are often known as ‘gate-keeper’ questions (33), in that the CIDI does not ask about any other symptoms of depression if both these core symptoms are absent.

The GHQ-12 does not include any core or gate-keeper questions, all twelve questions need to be answered. Self-report instruments, such as the GHQ-12, are therefore often able to provide more information on specific symptoms of depression than comprehensive instruments such as the CIDI (23), although the latter provides an overall diagnosis of depression based on DSM-IV criteria.

The GHQ-12 refers to the past “few weeks”, however current CIDI DSM-IV depression was defined as that having occurred in the past twelve months. Twelve-month diagnoses were therefore employed, consistent with the period of time generally employed in epidemiological studies using this instrument (4).

A twelve-month period was also consistent with the indicator of recency employed in the doctor diagnosed or treated medical conditions, as well as the other psychological disorders defined below. Previous studies have shown a relationship between past-month self-report symptoms and past-twelve-month depression (23).

5.2.3 Comorbid physical and psychological illness

Physical illness or medical conditions were assessed using the following criteria:

- a case definition of medically unexplained chronic fatigue based upon the collaborative International Chronic Fatigue Syndrome Study Group (34) criteria. Briefly, medically unexplained chronic fatigue is defined as the presence of at least six months of extreme tiredness or fatigue for which no explanatory medical or psychological condition, such as major depression, was identifiable in the previous twelve months (35). The presence of depressive symptoms such as feelings of worthlessness did not preclude the presence of medically unexplained chronic fatigue, unless a more severe defined major depressive disorder based on DSM-IV criteria was present.
- a case definition of multisymptom illness which was based upon the Centers for Disease Control and Prevention operational definition (36), and adopted for studies of United States (37), United Kingdom (38) and Australian Gulf War veterans (10).
- Self-reported doctor diagnosed or treated high blood pressure, asthma, (39) and any musculoskeletal disorder (40).

At the medical assessment, the above doctor diagnosed medical conditions were reviewed by medical doctors as to the likelihood of the diagnosis of these conditions, employing a variety of specific criteria (39, 41). To be included in the current statistical analyses, the physician assessment of the likelihood of medical diagnosis of the condition must have been classified as ‘probable’, and the condition have been treated by a doctor in the past twelve months, as an indicator of currency (40). Assessments regarded as possible, unlikely or non-medical were

treated as absent, as were self-reports of absence of that condition (41). Any musculoskeletal disorder was defined as the presence of one or more of arthritis or rheumatism, back or neck problems, joint problems, and/or soft tissue disorders (40).

Psychological disorders were assessed using CIDI past twelve-month diagnoses of anxiety disorder (including PTSD), PTSD alone, and alcohol use disorders (11).

Health-related quality of life was assessed using the twelve item Short-Form Health Survey (SF-12), version one, a widely employed measure of well-being and health-related quality of life (12, 14). The SF-12 consists of two summary scales, the Physical Component Summary (PCS-12), an index of physical health and well-being, and the Mental Health Component Summary (MCS-12), an index of mental health and well-being. Both summary scales use the same items, as shown in Appendix 3, but are weighted differently when scored (42).

5.3 Statistical analyses

Initial descriptive analyses were carried out to provide summary measures, such as frequencies, for the individual GHQ-12 items. Logistic regression was employed to assess the relationship between twelve-month CIDI DSM-IV major depression and each GHQ-12 item individually (using twelve separate regressions) and all twelve GHQ-12 items simultaneously (18, 23). Identification of key GHQ-12 items and their combinations was performed using several variable selection methods (43) comprising stepwise regression, decision tree and decision rule techniques, as follows.

5.3.1 Stepwise logistic regression

Stepwise regression (44) is an automated computational procedure that attempts to search for a subset of variables from a larger set of variables based on their association with the outcome variable. In this study stepwise regression was used to select a subset of GHQ-12 items based on their association with major depression. At each step of the process, either the best-performing variable, in terms of its association with the outcome is added to the regression model (forward entry), or the worst-performing variable is dropped from the model (backward elimination) (45). Stepwise regression has recently been applied to brief screening test construction (21, 46).

5.3.2 Decision trees

Decision trees such as Classification and Regression Trees (CART) (47), seek to identify subgroups, defined by combinations of variables such as depressive symptoms, with similar characteristics on an outcome variable, such as major depression. Unlike most implementations of stepwise regression, CART explicitly validates the generality of its models, using by default ten-fold cross-validation (47). Datasets are randomly divided into ten subsets, each subset in turn being used to test the performance of the tree created with the other nine subsets. CART and other tree-building methods have been applied in a wide variety of epidemiological and other relevant studies (18, 48).

5.3.3 Decision rules

CART provides a readily interpretable decision tree, graphically representing associations between depressive symptoms and depression. A possible disadvantage of decision trees, however, is that certain types of logical relationships between items or symptoms cannot easily be represented. In particular, decision trees may have trouble detecting relationships involving both logical *or*'s (screening rule true if at least one specified item is present, e.g., items A *or* B *or* C are present) as well as logical *and*'s (screening rule true if *all* specified items are present, e.g. items A *and* B *and* C are present) (49). For example, a decision tree cannot readily generate a screening rule such as:

Depression is present if (items A *or* B) *and* (items C *or* D) are present.

Decision rules are similar to decision trees, except that they are more flexible in terms of combining items, and are able to generate screening rules such as the one shown above. Although decision rules are not as widely used as decision trees, depression screening rules of the above type have been previously obtained by members of the research team (22), using the SPAN (Search Partition Analysis) decision rule technique (49, 50). SPAN employs a similar cross-validation technique to CART.

5.3.4 Threshold or cut-off score

Finally, as optimum thresholds or cut-off scores for GHQ-12 have been found to vary across different populations (12, 51), we empirically established cut-off GHQ-12 scores for the

presence of CIDI DSM-IV major depression using standard Receiver Operating Characteristic (ROC) curve techniques (52). ROC analysis attempts to determine a balance between sensitivity and specificity, as defined below.

5.3.5 Screening and diagnostic performance of the different methods

The screening and diagnostic performance of the GHQ-12 items chosen by the above techniques, in regard to diagnosing twelve-month major depression, was assessed using standard measures including:

- sensitivity (true positive rate, or the number of veterans with CIDI DSM-IV major depression correctly classified by the depression screening test)
- specificity (true negative rate, or the number of veterans without major depression correctly classified by the screening test)
- positive predictive value (the number of veterans testing positive on the screening test actually having been diagnosed with major depression)
- negative predictive value (the number of veterans testing negative on the screening test who were not diagnosed with major depression)
- area under the ROC curve (AUC). The latter is equal to the probability that a random person meeting diagnostic criteria (e.g., major depression) has a higher value of the measurement (e.g. a positive test result on the depression screen) than a random person not meeting diagnostic criteria. An area under the curve of 0.50 denotes chance level performance (52). In order to compare the overall performance between the various screening methods, areas under ROC curves were statistically compared (53).

5.3.6 Comorbid physical and psychological illness

The association between a depression screen based on the key combination of GHQ-12 items identified above and the physical, psychological and health-related quality of life measures was ascertained using logistic regression and median regression methods. Logistic regression was employed to compare those veterans scoring present on a GHQ-12 depression screening test, with those scoring absent on that screening test, on each binary outcome variable representing comorbid psychological or physical illness. The presence or absence of an

affirmative result on the depression screening test, suggesting possible depression caseness, in combination with the presence or absence of one or more physical or psychological conditions (consisting of alcohol use disorder and or PTSD, anxiety disorders being dropped to avoid redundancy), was then compared on SF-12 mental and physical health-related quality of life.

Tests of effect modification or statistical interaction (54) were employed to ascertain whether the difference in quality of life between the presence and absence of any comorbid physical or psychological disorder, was increased by having an affirmative result on the depression screen. In other words, statistical interaction analysis tests whether the presence of both possible depression caseness and at least one comorbid physical or other psychological disorder combined leads to poorer quality of life than either depression caseness by itself or any physical or other psychological disorder by itself.

Median regression (55) was employed for the analysis of the SF-12, as prior investigation of the Australian Gulf War Veterans' Health Study data found this measure to exhibit a small number of extremely low scores, representing poor physical or mental health (40). Median regression is less affected by extreme scores than is conventional linear regression.

Logistic and median regression were undertaken separately for the depression screen based upon key GHQ-12 symptoms, as well as the conventional screening cut-off or threshold based upon the total GHQ-12 score. This comparison was performed in order to determine whether a brief screening instrument based upon key GHQ-12 items was as closely associated with physical and psychological comorbidity and quality of life, as was the screening cut-off based upon the total GHQ-12 score.

5.3.7 Possible confounding factors

All logistic and median regression analyses, apart from those involving stepwise regressions, were adjusted for the effects of possible confounding factors, comprising age category at time of Gulf War study (less than 20 years old, greater than 20 but less than or equal to 25 years, greater than 25 but less than or equal to 35 years, greater than 35 years old), military rank at time of 1990/91 Gulf War (other ranks, non-supervisory; other ranks, supervisory; officers), smoking status (current, former, never), and obesity. Obesity was defined as a body mass index (BMI) equal to or greater than 30 kilograms/metres² (56). The results of the crude and adjusted odds ratios (OR) and regression coefficients (representing the differences between

medians) were very similar, and so only the adjusted ORs and median regression coefficients are provided throughout, except in the case of stepwise regressions.

The stepwise regressions were not adjusted for possible confounding factors so as to maintain parity with the decision tree and decision rule techniques, in which it is more difficult to include confounding factors. The primary purpose of these three approaches was to identify key GHQ-12 items, and so only crude or unadjusted ORs are presented for the stepwise regressions.

5.3.8 Statistical software

General statistical analyses, such as logistic regression and median regression were performed using the Stata statistical package, version 12 (Stata Corporation, College Station, Texas, 2011).

CART analyses were conducted using CART software (Salford Systems, San Diego, California, 2012). SPAN analyses were conducted using SPAN software (Department of Community Health, University of Auckland, New Zealand, 2011).

6 Results

6.1 Individual and simultaneous relationships between GHQ-12 items and major depression

Numbers and percentages of male RAN Gulf War veterans with each individual GHQ-12 item present, as well as the results of the individual logistic regressions on each item are provided in Table 1. It can be seen in Table 1 that GHQ-12 items five “Felt constantly under strain”, nine “Feeling unhappy and depressed” and two “Lost much sleep over worry” were the most frequently occurring items. Items eleven “Been thinking of yourself as a worthless person” and four “(not) felt capable of making decisions” were the least frequently occurring items.

All twelve individual GHQ-12 items were significantly associated with CIDI DSM-IV depression, with the highest odds ratio being observed for items eleven “Been thinking of yourself as a worthless person”, four “(not) felt capable of making decisions” and eight “(not) been able to face up to your problems”.

Table 1 Logistic regression of 12-month CIDI DSM-IV major depression by individual GHQ-12 items

Item	Label	N	%	adjusted OR (95% CI)
1	(not) been able to concentrate on whatever you are doing	281	23.4	5.55 (3.63-8.48)
2	lost much sleep over worry	300	25.0	5.56 (3.64-8.49)
3	(not) felt playing useful part in things	226	18.8	5.29 (3.46-8.10)
4	(not) felt capable of making decisions about things	115	9.6	9.87 (6.08-16.03)
5	felt constantly under strain	396	32.9	4.68 (3.04-7.19)
6	felt couldn't overcome your difficulties	179	14.9	7.85 (5.06-12.18)
7	(not) been able to enjoy your normal day to day activities	283	23.6	5.74 (3.74-8.80)
8	(not) felt able to face up to your problems	125	10.4	7.92 (5.00-12.62)
9	been feeling unhappy and depressed	321	26.7	6.44 (4.17-9.93)
10	been losing confidence in yourself	189	15.7	6.27 (4.07-9.66)
11	been thinking of yourself as a worthless person	88	7.3	13.17 (7.84-22.15)
12	(not) been feeling reasonably happy, all things considered	210	17.5	8.05 (5.19-12.47)

*Odds ratios (ORs), obtained using separate logistic regressions for each GHQ-12 item individually, are adjusted for the possible confounders described in the report

It can be seen in Table 2, that when all of the GHQ-12 items were entered into a logistic regression model simultaneously, only item eleven, pertaining to feelings of worthlessness was statistically significant, possibly as a result of inter-correlation between the symptoms (18, 23).

Table 2 Logistic regression of 12-month CIDI DSM-IV major depression by all GHQ-12 items entered simultaneously

Item	Label	adjusted OR* (95% CI)
1	(not) been able to concentrate on whatever you are doing	1.43 (0.79-2.58)
2	lost much sleep over worry	1.59 (0.89-2.84)
3	(not) felt playing useful part in things	1.31 (0.79-2.58)
4	(not) felt capable of making decisions about things	1.73 (0.83-3.61)
5	felt constantly under strain	1.17 (0.64-2.11)
6	felt couldn't overcome your difficulties	1.70 (0.86-3.34)
7	(not) been able to enjoy your normal day to day activities	1.15 (0.61-2.19)
8	(not) felt able to face up to your problems	1.00 (0.49-2.05)
9	been feeling unhappy and depressed	1.63 (0.83-3.22)
10	been losing confidence in yourself	0.72 (0.36-1.46)
11	been thinking of yourself as a worthless person	2.93 (1.41-6.07)
12	(not) been feeling reasonably happy, all things considered	1.56 (0.79-3.07)

*ORs , obtained using a single logistic regression model for all GHQ-12 items simultaneously, are adjusted for the possible confounders described in the report

Crude ORs were obtained for stepwise regression, and are provided in Table 3. The same five GHQ-12 items were chosen by both forward entry and backward elimination stepwise methods defined earlier, with the largest odds ratio (2.93) being observed for GHQ-12 item eleven, "Been thinking of yourself as a worthless person".

Table 3 Stepwise Logistic regression of 12-month CIDI DSM-IV major depression by GHQ-12 items chosen by both forward and backward methods

Item	Label	OR* (95% CI)
2	lost much sleep over worry	1.80 (1.10-3.08)
4	(not) felt capable of making decisions	2.10 (1.14-3.84)
6	felt you couldn't overcome your difficulties	2.01 (1.11-3.62)
9	been feeling unhappy and depressed	2.12 (1.20-3.76)
11	been thinking of yourself as a worthless person	3.07 (1.64-5.73)

*Crude ORs, obtained using both forward and backward stepwise logistic regression methods, described in the text. Note that confidence intervals do not take into account the fact that the variables were selected using stepwise regression, which is an automated search procedure

6.2 CART decision tree and SPAN decision rule for key GHQ-12 items

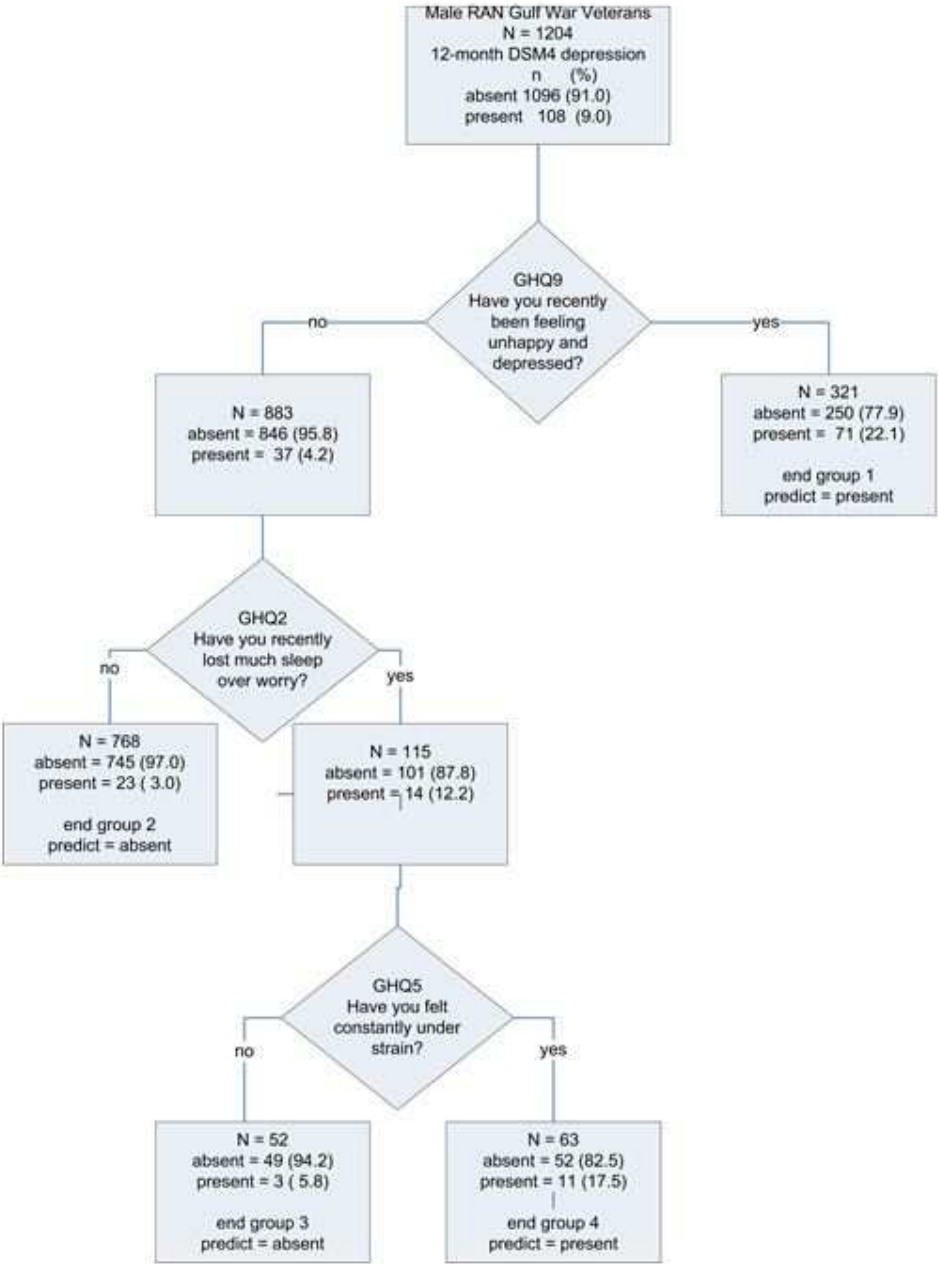
CART generated the decision tree shown in Figure 1. At each stage of the tree-growing process, CART chooses that variable with the lowest Gini index. The Gini index provides a measure of subgroup homogeneity or purity, on which to split the dataset (47). Splitting is continued until each subgroup, or “end group” is either fully homogenous (only one outcome present), or becomes too small (e.g., five observations or less) for further splitting. The resulting tree is then pruned back, balancing model “cost” (cross-validated performance) against model “complexity” (number of subgroups) (47).

The variable first selected by CART to split the dataset was GHQ-12 item nine “Been feeling unhappy and depressed”. The total dataset was therefore split into two subgroups, those that scored absent on GHQ-12 item nine, and those that scored present on this item. The subgroup consisting of veterans that scored present on item nine was not able to be split any further, as any increase in the number of subgroups was not accompanied by a commensurate increase in performance. This node is therefore classed as an end group (end group one in Figure 1).

The subgroup that responded absent to item nine was subsequently split on GHQ-12 item two “Lost much sleep over worry”. The subgroup of veterans responding absent to item two was

not able to be split any further, and so forms the second end group. The subgroup of veterans responding present to item two were then split on item five "Felt constantly under strain". Neither of the resulting subgroups was able to be split any further, forming end groups three and four, corresponding to those veterans who responded no and yes to item five respectively.

Figure 1 CART decision tree for CIDI DSM-IV Major Depression, identifying key GHQ-12 items

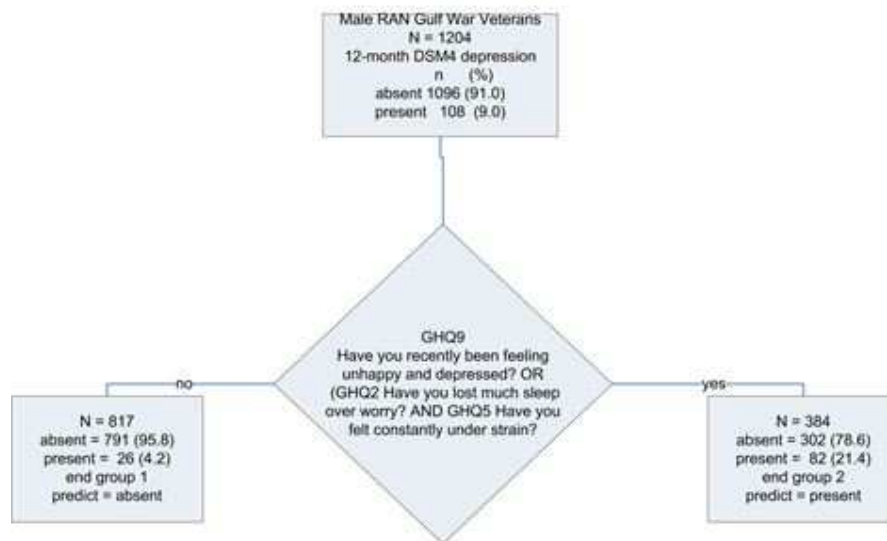


For end groups one and four CART classified major depression as being present, while for end groups two and three, CART predicted major depression as being absent. The decision tree described above involves three GHQ-12 items and four end groups. In order to ascertain

whether the above decision tree could be simplified, a variable representing the four end groups was itself entered into CART, in a process known as subgroup amalgamation (57). The four CART subgroups were subsequently merged into two subgroups, as shown in Figure 2.

If GHQ-12 item nine “Been feeling unhappy and depressed” was present, or if item nine was absent but both items two “Lost much sleep over worry” and five “Felt constantly under strain” were present, CART classified major depression as being present. If the above rule was not true, that is if item nine was absent, and either or both of items two and five were absent, then CART classified major depression as being absent.

Figure 2 Merged CART decision tree for CIDI DSM-IV Major Depression, identifying key GHQ-12 items



As mentioned above, SPAN decision rules can often more easily link items using logical *and*'s and *or*'s than can CART. SPAN generated a very simple decision rule, consisting of only two items as shown below.

Possible depression is present if GHQ-12 item nine "Feeling unhappy and depressed" is present, *or* if item two "Lost much sleep over worry" is present.

The above SPAN decision rule is true if either one of the items is present, or if both items are present. The two GHQ-12 items chosen by SPAN (items nine and two) were among the three items chosen by CART.

6.3 Twelve item cut-off / threshold for GHQ-12 total score

The recommended GHQ-12 screening cut-off or threshold for determining possible caseness for general psychological distress, which the GHQ was originally designed to measure, is a total score of three or more (24). We obtained this same threshold empirically in the present study, in regard to the presence of CIDI twelve-month DSM-IV major depression.

6.4 Screening and diagnostic performance of the above methods

The screening and diagnostic performance of the instrument based upon all twelve GHQ-12 items (GHQ-12 screening cut-off or threshold), as well as the brief screening instruments (stepwise logistic regression, CART decision tree and SPAN decision rule), is provided in Table 4. The performance of a simultaneous logistic regression on all twelve GHQ-12 items is also shown for purposes of comparison.

It can be seen in Table 4 that the simultaneous logistic regression, with twelve GHQ-12 items, and the stepwise logistic regression, with five GHQ-12 items, exhibit slightly higher AUC's (Area under the Receiver Operating Characteristic Curve) than the screening instruments generated using the other methods, which all have similar AUC's. AUC was significantly higher for the simultaneous logistic regression model (AUC = 0.82, 95% CI = 0.78 to 0.87) than for the CART instrument (AUC = 0.74, 95% CI = 0.70 to 0.78), $p < 0.001$. The former technique outputs a continuous probability of group membership (e.g., depression) whereas CART outputs a simple binary classification as to the presence or absence of depression.

Logistic regression will therefore tend to exhibit higher areas under ROC curves than will tree-building techniques such as CART (23, 48).

All instruments exhibited very similar performance on the other screening and diagnostic measures, except that the SPAN decision rule had a lower specificity than the other instruments. The CART tree was the second-briefest instrument generated, yet had similar performance to the other instruments, although it had higher specificity than the smallest instrument, the two item SPAN decision rule.

As the three item CART tree appeared to achieve the best balance between number of items, and performance, it was chosen as the brief depression screening instrument, representing a key combination of GHQ-12 items.

The three item CART brief depression screen classified possible depression as being present if the veteran was feeling “unhappy and depressed”, or, in the absence of this symptom, if the veteran had both “lost much sleep over worry” and “felt constantly under strain”.

Table 4 Screening and diagnostic performance of screening rules generated by each statistical or computational method

	Simultaneous logistic regression	Stepwise logistic regression	CART decision tree	SPAN decision rule	Cut-off (3+)
Number of items	12	5	3	2	12
Sensitivity (95% CI)	0.76 (0.67 – 0.84)	0.73 (0.64 – 0.81)	0.76 (0.67 – 0.84)	0.79 (0.70 – 0.86)	0.77 (0.68 – 0.84)
Specificity (95% CI)	0.75 (0.72 – 0.77)	0.72 (0.70 – 0.75)	0.72 (0.70 – 0.75)	0.68 (0.65 – 0.71)	0.73 (0.70 – 0.75)
PPV (95% CI)*	0.23 (0.18 – 0.27)	0.21 (0.17 – 0.25)	0.21 (0.17 – 0.26)	0.19 (0.16 – 0.24)	0.22 (0.18 – 0.26)
NPV (95% CI)[†]	0.97 (0.96 – 0.98)	0.96 (0.95 – 0.98)	0.97 (0.95 – 0.98)	0.97 (0.96 – 0.98)	0.97 (0.96 – 0.98)
AUC (95% CI)[‡]	0.82 (0.78 – 0.87)	0.80 (0.75 – 0.84)	0.74 (0.70 – 0.78)	0.73 (0.69 – 0.77)	0.75 (0.71 – 0.79)

*PPV = positive predictive value

[†]NPV = negative predictive value

[‡]AUC = area under the receiver operating characteristic curve; definitions provided in the text

6.5 Brief depression screen and medical and psychological conditions

The associations between the CART brief depression screen, and medical and psychological conditions were then examined. The associations between the above variables and the cut-off for the total GHQ-12 score were also examined. The latter step was performed in order to ascertain whether the cut-off score and the brief screening instrument exhibited similar associations with comorbid medical and psychological conditions. Table 5 shows that male RAN Gulf War veterans with the three item CART brief depression screen present were more likely to exhibit each of the medical or psychological conditions than those with this screen absent.

For example, of the 812 veterans scoring absent on the CART brief depression screen, 52 (6.4%) had doctor-diagnosed high blood pressure. Of the 383 veterans scoring present on the brief depression screen, 33 (8.6%) had doctor-diagnosed high blood pressure.

The largest differences in the risk of medical or psychological conditions between those veterans with the brief depression screen absent, and those with the screen present were observed for PTSD, multisymptom illness and chronic fatigue. All differences in relation to the presence or absence of the CART depression screen and the presence or absence of physical or psychological conditions were statistically significant with the exception of high blood pressure.

Table 5 Relationship between presence of CART brief depression screen and medical and psychological conditions

		Number (%) of veterans with the outcome in each group*			
Outcome		Brief depression screen absent		Brief depression screen present	
Medical or psychological condition	N	n (%)	N	n (%)	adjusted OR [†] (95% CI)
High blood pressure	812	52 (6.4)	383	33 (8.6)	1.46 (0.92 – 2.36)
Asthma	812	85 (10.5)	383	57 (14.9)	1.52 (1.06 – 2.20)
Musculoskeletal disorders	792	163 (20.6)	369	112 (30.4)	1.74 (1.31 – 2.32)
Multisymptom illness	817	118 (14.4)	384	188 (49.0)	5.65 (4.25 – 7.55)
Chronic fatigue	815	31 (3.8)	383	61 (15.9)	4.65 (2.94 – 7.36)
Alcohol use disorder	817	21 (2.6)	384	33 (8.6)	3.56 (2.02 – 6.27)
Any anxiety disorder	817	62 (7.6)	384	94 (24.5)	3.89 (2.73 – 5.54)
PTSD	817	17 (2.1)	384	43 (11.2)	5.66 (3.16 – 10.12)
Any psychological disorder	817	104 (12.7)	384	150 (39.1)	4.37 (3.25 – 5.88)
Any physical or psychological condition	817	474 (58.0)	384	275 (71.6)	3.51 (2.69 – 4.56)

*For example, of the 812 veterans scoring absent on the CART brief depression screen, 52 (6.4%) veterans had doctor-diagnosed high blood pressure. Of the 383 veterans scoring present on the brief depression screen, 33 (8.6%) veterans had doctor-diagnosed high blood pressure

[†]ORs are adjusted for are adjusted for the possible confounders described in the text

It can be seen in Table 6 that similar results were obtained for associations between a total GHQ-12 score of equal to or greater than the cut-off of three versus a score less than three, and the presence of physical or psychological conditions. The associations between GHQ-12 cut-off score and both high blood pressure and asthma, however, were not statistically significant.

Table 6 Relationship between twelve item cut-off (GHQ-12 score => 3) and physical and psychological conditions

		Number (%) of veterans with the outcome in each group*			
Outcome		Total GHQ-12 score < 3		Total GHQ-12 score => 3	
Condition	N	n (%)	N	n (%)	adjusted OR [†] (95% CI)
High blood pressure	817	56 (6.8)	379	29 (7.7)	1.16 (0.72 – 1.86)
Asthma	817	88 (10.8)	379	54 (14.3)	1.41 (0.98 – 2.04)
Musculoskeletal disorders	797	164 (20.6)	365	111 (30.4)	1.74 (1.31 – 2.32)
Multisymptom illness	822	122 (14.8)	379	184 (48.4)	5.47 (4.11 – 7.29)
Chronic fatigue	820	33 (4.0)	380	59 (5.6)	4.34 (2.76 – 6.84)
Alcohol use disorder	822	19 (2.3)	380	35 (9.2)	4.28 (2.40 – 7.68)
Any anxiety disorder	822	58 (7.1)	380	98 (25.8)	4.56 (3.19 – 6.53)
PTSD	822	15 (1.8)	380	45 (11.8)	6.97 (3.81 – 12.73)
Any psychological disorder	822	97 (11.8)	380	157 (41.3)	5.33 (3.94 – 7.19)
Any physical or psychological disorder	822	355 (43.2)	380	263 (69.2)	2.98 (2.29 – 3.86)

*For example, of the 817 veterans whose total GHQ-12 score was less than the screening cut-off of 3, 56 (6.8%) veterans had doctor-diagnosed high blood pressure. Of the 379 veterans scoring present on the GHQ-12 depression cut-off (total GHQ-12 score of 3 or greater), 29 (7.7%) veterans had doctor-diagnosed high blood pressure

[†] ORs, obtained using logistic regression, are adjusted for the possible confounders described in the text

6.6 Comorbid CART brief depression screen and medical / psychological conditions and health-related quality of life

A comparison of the presence and absence of comorbidity of depression screen and physical / psychological conditions on health-related quality of life is shown in Table 7. The results of the median regression indicated that those veterans with both the CART brief depression screen

present, and any comorbid physical or psychological condition present, had the lowest median SF-12 MCS and PCS scores, reflecting lower SF-12 mental and physical health-related quality of life.

The difference in both SF-12 MCS and PCS scores between the presence and absence of a comorbid physical or psychological condition was greater for the CART brief depression screen present, than it was for the depression screen absent. In other words, mental (SF-12 MCS) and physical (SF-12 PCS) health-related quality of life were at their lowest levels if the brief depression screen and at least one comorbid physical or psychological condition were both present. Such a result suggests the presence of a statistical interaction between the brief depression screen and physical / psychological comorbidity in regard to health-related quality of life, which was in fact statistically significant ($p < 0.05$) for both SF-12 MCS and PCS scores.

Table 7 Comparison of comorbidity of CART brief depression screen and presence/absence of physical and psychological conditions on SF-12 mental and physical health summary scales

		CART brief depression screen				
		Present		Absent		
	Any other physical or psychological condition	Median (IQR [†])	adjusted Regression coefficient (95% CI) ^{†,‡}	Median (IQR*)	adjusted Regression coefficient (95% CI) ^{†,‡}	P value for test of interaction between CART brief depression screen and any other Physical or Psychological Condition
SF-12 MCS	Absent	42.05 (16.22)	-5.60 (-8.74 – -2.46)	55.04 (7.11)	-1.67 (-3.07 – -0.27) [§]	0.021
	Present	35.68 (14.93)		53.14 (11.55)		
SF-12 PCS	Absent	54.71 (6.97)	-8.89 (-11.51 – -6.27)	54.21 (5.00)	-4.76 (-6.23 – -3.29)	0.019
	Present	46.50 (16.60)		49.60 (14.00)		

*Inter-quartile range (difference between 75th and 25th percentiles)

[†] Median regression, confidence intervals obtained using 10,000 bootstrap replications

[‡] Median regression coefficients are adjusted for the possible confounders described in the text

[§] Test of statistical interaction between CART GHQ-12 screen and presence of any other physical or psychological condition on median SF-12 MCS values, $p = 0.021$

^{||} Test of statistical interaction between CART GHQ-12 screen and presence of any other physical or psychological condition on median SF-12 PCS values, $p = 0.019$

Similar results were obtained in regard to the combination of twelve item depression cut-off and any other physical or psychological comorbidity, except that the statistical interaction between depression cut-off and comorbidity for physical health-related quality of life was more pronounced ($p < .001$), as shown in Table 8.

Table 8 Comparison of comorbidity of twelve item cut-off and presence/absence of physical and psychological conditions on SF-12 mental and physical health summary scales

	GHQ-12 total score					P value for test of interaction between GHQ-12 Depression Screen and Any Other Physical or Psychological Condition
		Score =>3		Score <3		
	Any other Physical or Psychological Condition	Median (IQR*)	adjusted Regression coefficient (95% CI) ^{†,‡}	Median (IQR)	adjusted Regression coefficient (95% CI) ^{†,‡}	
SF12-MCS	Absent	42.05 (16.22)	-4.38 (-9.22 - -3.51)	55.04 (7.11)	-2.24 (-3.48 - -0.999) [§]	0.049
	Present	35.68 (14.93)		53.14 (11.56)		
SF-12 PCS	Absent	54.71 (6.97)	-10.30 (-13.02 - -7.60)	54.21 (5.00)	-4.21 (-5.38 - -3.04)	< 0.001
	Present	46.50 (16.60)		49.60 (14.00)		

*Inter-quartile range (difference between 75th and 25th percentiles)

[†] Median regression, confidence intervals obtained using 10,000 bootstrap replications

[‡] Median regression coefficients are adjusted for the possible confounders described in the text

[§] Test of statistical interaction between cut-off and presence of any other physical or psychological condition on median SF-12 MCS values, $p = 0.049$

^{||} Test of statistical interaction between cut-off and presence of any other physical or psychological condition on median SF-12 PCS values, $p < 0.001$

7 Discussion

7.1 Key depressive symptoms

A brief screening instrument for depression was identified using the CART tree-building technique, which performed similarly to longer screening instruments in terms of correctly classifying those veterans with a CIDI DSM-IV diagnosis of major depression. The brief instrument consisted of three items, “been feeling unhappy and depressed” by itself and the combination of “lost much sleep over worry” and “felt constantly under strain”.

In regard to the above symptoms, depressed mood, and sleeping either less or more than usual, are part of DSM-IV (and DSM-5 (31)) criteria for major depression. Along with anhedonia or loss of interest or pleasure, depressed mood is a core DSM-IV symptom. At least one of these two ‘gatekeeper’ symptoms needs to be present for a DSM-IV diagnosis of major depression to take place.

Feeling constant strain is not part of the DSM-IV criteria for major depression, however symptoms such as feelings of pessimism that are not part of the formal DSM-IV criteria for major depression can nevertheless be highly associated with a DSM-IV diagnosis of major depression (18).

Although depressed mood is a core DSM symptom, the individual logistic regression analyses suggested that depressed mood was not the GHQ-12 item most highly associated with twelve-month CIDI DSM-IV major depression. The highest individual associations were actually observed for “been thinking of yourself as a worthless person”, and “(not) capable of making decisions”. Both these GHQ-12 items exhibited low prevalences ($n = 88$, 7.3% and $n = 115$, 9.6% respectively), reflected in the broad confidence intervals, which is suggestive of why they were not selected by the CART procedure.

In regards to the general population, sleep disturbance has been shown to be a strong marker of major depression, even amongst those with medical comorbidity (58), although it should be noted that the actual GHQ-12 item specifies lack of sleep due to worry, rather than due to a nonspecific cause. Sleep disturbance, including insomnia and nightmares, is a major problem in veteran populations (59). Sleep disturbance in veterans has recently been found to be symptomatic of major depression (59), in addition to being associated with decreased overall functioning (60).

Strain can be defined as “those enduring problems that have the potential for arousing threats” (p.3) (61). Job strain, often characterised as a sense of high demands but low feelings of control, is associated with PTSD in veterans (62), although more research is needed to obtain more details about what ‘feelings of strain’ are actually associated with.

7.2 Screening and diagnostic performance of brief depression screen

Previous results regarding the performance of brief screening tests have been mixed. Many studies have found brief instruments to perform similarly to the longer instruments from which they were derived (23, 25, 63), although other studies have found brief instruments to exhibit poorer performance (46, 64). With regards to studies specifically based upon the GHQ, albeit the twenty-eight item version (24), four item instruments, obtained using decision rule methods (22, 65), performed as well as did cut-offs based upon the longer instruments.

Although further testing on veteran and military populations is necessary to ascertain generality, our screening instrument exhibited similar screening and diagnostic performance to instruments based upon logistic regression models and cut-off scores involving all twelve GHQ-12 items.

7.3 Brief depression screen and physical and psychological conditions

Scoring present on the brief depression screen was associated with several physical and psychological conditions, consisting of asthma, musculoskeletal disorders, multisymptom illness, chronic fatigue, alcohol use disorder, any anxiety disorder, PTSD, any psychological disorder, and any physical or psychological disorder. There was no statistically significant association between presence of the brief depression screen and the presence of high blood pressure. Similar findings were observed in the case of the GHQ-12 cut-off, although in addition to high blood pressure there was no statistically significant association with asthma.

Although our finding regarding a lack of association between possible depression caseness and high blood pressure is contrary to that of general population studies (66), our other

findings regarding physical and psychological disorders are in agreement with such studies (67).

In our previously published papers we have established associations between multisymptom illness (10), musculoskeletal disorders (40), anxiety disorders and alcohol disorders (11), with DSM-IV major depression. The present results extend these associations to a brief self-report measure of depression.

The presence of psychiatric disorders such as depression can interfere with the recognition and treatment of physical disorders, and vice versa (68). The management of comorbid physical and psychological conditions has recently been identified as an especially challenging area for both patients and health professionals, partly due to the perceived stigma of the latter conditions (69).

7.4 Comorbid brief depression screen and physical / psychological conditions and health-related quality of life

Veterans scoring present on the brief depression screen, and with at least one comorbid physical or psychological condition, exhibited lower mental and physical health and well-being than those with either the brief depression screen, or a comorbid condition, by itself. Similar results were obtained for the GHQ-12 cut-off.

The relationship between physical and psychological disorders appears to operate in both directions (69). Physical illness such as musculoskeletal disorders and asthma may impact on the ability to perform everyday work, social and home activities for example, increasing the likelihood of depression (69).

7.5 Implications for identification of depression in veterans

As discussed earlier, depression is very difficult to recognise within general medical or community settings. Lack of identification may be due to a variety of reasons, such as masking of the symptoms of depression, particularly physical or somatic ones such as

sleeping and eating disturbances, by other disorders, and time constraints within a GP or medical practitioner visit (70).

The brief depression screen consists of only three items, while the sequential nature of decision trees (23) means that if depressed mood is present, no further questions need be asked. If depressed mood is absent, however, information on the presence or absence of the other two symptoms would be required. One of the latter two symptoms is a physical one, lack of sleep, but it is specified as lack of sleep due to worry, and must occur in combination with feeling constantly under strain.

It is thought that the above three symptoms could be recognised by family members and medical practitioners, as well as veterans themselves. To facilitate such recognition, a simple three item rule such as the CART brief depression screen could be simply printed on a wallet-sized card.

A brief depression screen could also be readily implemented on a website (71). Such a screen could be augmented by online information about depression, or online exercises intended to teach the basic principles of cognitive behaviour therapy or other best practice treatments for depression, similar to the Australian National University MoodGYM self-help online program (72) (www.moodgym.anu.edu.au).

No matter how brief a screening instrument is, depression and other psychological disorders still carry stigma, particularly in the serving military (73). It is hoped that education programs may counter such stigma, although in the interim the specific choice of wording of brief questionnaires may be important. It has been observed for example, that “down-hearted and blue” was found to exhibit better screening properties amongst hospital inpatients than “depressed”, possibly due to the negative connotations of the latter (74). Future research may also need to look at the exact wording of the “Have you recently been feeling unhappy and depressed” item of the brief depression screen.

It must also be emphasised that brief depression screens are intended for screening for, or detecting, possible cases of depression, which then need to be followed up by appropriate health professionals, a screening test obviously not being an actual diagnostic instrument.

Given the association between depression and physical and psychological disorders, and the adverse effects of comorbid physical and psychological illness on health-related quality of life,

online information, as well as more formal treatment programs, could form part of a wellness approach to physical and mental health. Comorbid physical and psychological illnesses would ideally be treated concurrently rather than sequentially (75). For example, a combined antidepressant therapy and pain self-management program has been found to be effective in primary care patients with comorbid depression and musculoskeletal pain (76).

Although not intended as a comprehensive comparison of screening techniques, our results suggest that when constructing brief screening tests, both CART decision tree and SPAN decision rule approaches should be examined, and compared with the longer established (65), but not necessarily more effective, cut-off approach.

7.6 Implications for further research

It is important that the performance of the three item brief depression screen be tested on other populations of veterans to ascertain its generality. A useful initial analysis would be to examine the performance of the brief depression screen on the Australian Gulf War Veterans' Health Study Follow-Up dataset. This would allow the relationship between baseline GHQ-12 items and CIDI DSM-IV major depression diagnosed at follow-up, ten years later, to be ascertained.

Subsequent analysis of the follow-up dataset would also enable us to determine how well the three items selected for the baseline dataset perform when applied to the follow-up dataset. The performance of the brief depression screen with one based upon identifying the key items of the PHQ-9, the nine item depression scale of the Patient Health Questionnaire (77), could also be examined.

The PHQ-9 consists of the nine DSM-IV symptoms of major depression, outlined in Appendix 2, but provides a dimensional rather than categorical (yes / no) measure of depression. Each PHQ-9 symptom is coded on a four point scale (e.g., 'not at all', 'several days', 'more than half the days' and 'nearly every day', in regards to how often they were bothered by a symptom such as depressed mood in the past two weeks).

It should be noted, however, that the PHQ-9 was not administered during Gulf War Veterans' baseline study, only during the follow-up study.

Finally, it would also be of interest in future research to compare the performance of the three item brief depression screen with one based upon the Kessler Psychological Distress Scale (K10) (78). The latter is being increasingly employed in population studies, such as the Australian National Survey of Mental Health and Well-being (79), as well as for Internet-based screening for depression (71). There is a dearth of studies on brief versions of the K10, apart from the six item version (K6) (78). The K10 and K6, however, were not employed in the baseline or follow-up components of the Australian Gulf War Veterans' Health study.

7.7 Conclusion

Depression is a major and debilitating disorder that commonly occurs yet can be difficult to identify, particularly in general settings such as GP clinics. A three item brief depression screen, consisting of GHQ-12 items concerned with depressed mood, lack of sleep due to worry, and feelings of being under strain, performed similarly to a cut-off approach based upon the total GHQ-12 score in terms of association with DSM-IV major depression. The brief screen also performed comparably to the twelve item cut-off in terms of its association with physical and psychological comorbidity and health-related quality of life.

The brief depression screen needs to be further tested on other veteran and military populations but could readily be used by veterans, their families and medical practitioners in order to help identify the early stages of depression.

8 Appendices

8.1 Appendix 1: 12 item General Health Questionnaire (GHQ-12)

We would like to know how you have been feeling over the past few weeks.

Please tick the box that most closely describes your experience for each question

1. Have you recently been able to concentrate on whatever you're doing?

- ☐ Better than usual ☐ Same as usual ☐ Less than usual ☐ Much less than usual

2. Have you recently lost much sleep over worry?

- ☐ Not at all ☐ No more than usual ☐ Rather more than usual ☐ Much more than usual

3. Have you recently felt that you are playing a useful part in things?

- ☐ More so than usual ☐ Same as usual ☐ Less useful than usual ☐ Much less useful

4. Have you recently felt capable of making decisions about things?

- ☐ More so than usual ☐ Same as usual ☐ Less so than usual ☐ Much less capable

5. Have you recently felt constantly under strain?

- ☐ Not at all ☐ No more than usual ☐ Rather more than usual ☐ Much more than usual

6. Have you recently felt you couldn't overcome your difficulties?

- ☐ Not at all ☐ No more than usual ☐ Rather more than usual ☐ Much more than usual

7. Have you recently been able to enjoy your normal day-to-day activities?

- ☐ More so than usual ☐ Same as usual ☐ Less so than usual ☐ Much less than usual

8. Have you recently been able to face up to your problems?

- ☐ More so than usual ☐ Same as usual ☐ Less able than usual ☐ Much less able

9. Have you recently been feeling unhappy and depressed?

- ☐ Not at all ☐ No more than usual ☐ Rather more than usual ☐ Much more than usual

10. Have you recently been losing confidence in yourself?

- ☐ Not at all ☐ No more than usual ☐ Rather more than usual ☐ Much more than usual

11. Have you recently been thinking of yourself as a worthless person?

- ☐ Not at all ☐ No more than usual ☐ Rather more than usual ☐ Much more than usual

12. Have you recently been feeling reasonably happy, all things considered?

- ☐ More so than usual ☐ Same as usual ☐ Less able than usual ☐ Much less able

8.2 Appendix 2: DSM-IV (Diagnostic and Statistical Manual, version 4) criteria for major depression

Five or more of the following symptoms must have been present during the same 2 week period. At least one of the symptoms must be depressed mood or loss of interest or pleasure, must be present.

- (1) Depressed mood most of the day, nearly every day
- (2) Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day
- (3) Significant weight loss when not dieting or weight gain (change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day
- (4) Insomnia or hypersomnia [sleeping too much] nearly every day
- (5) Psychomotor agitation or retardation [slowing] nearly every day
- (6) Fatigue or loss of energy nearly every day
- (7) Feelings of worthlessness or excessive or inappropriate guilt nearly every day
- (8) Diminished ability to think or concentrate, or indecisiveness, nearly every day
- (9) Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide

8.3 Appendix 3: 12 item Short-Form Health Survey (SF-12)

1. *In general, would you say your health is:*

☐ Excellent ☐ Very good ☐ Good ☐ Fair ☐ Poor

The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

- | | Yes,
Limited a
Lot | Yes, Limited
a Little | No,
Not
Limited
at
All |
|--|--------------------------|--------------------------|------------------------------------|
| 2. Moderate activities , such as moving a table, pushing a vacuum cleaner, bowling, or playing golf | ☐ | ☐ | ☐ |
| 3. Climbing several flights of stairs | ☐ | ☐ | ☐ |

During the **past 4 weeks**, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

4. **Accomplished less** than you would like ☐ YES ☐ NO
5. Were limited in the **kind** of work or other activities ☐ YES ☐ NO

During the **past 4 weeks**, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

6. **Accomplished less** than you would like ☐ YES ☐ NO
7. Didn't do work or other activities as **carefully** as usual ☐ YES ☐ NO

8. During the **past 4 weeks**, how much did **pain** interfere with your normal work (including both work outside the home and housework)?

☐ Not at all ☐ A little bit ☐ Moderately ☐ Quite a bit ☐ Extremely

These questions are about how you feel and how things have been with you during the **past 4 weeks**. For each question, please give the one answer that comes closest to the way you have been feeling.

How much of the time during the **past 4 weeks**.

All of the Most of A Good Some of A Little of None of
Time the Time Bit of the the Time the Time the Time

Time

9. Have you felt calm and peaceful? ☐ ☐ ☐ ☐ ☐ ☐
10. Did you have a lot of energy? ☐ ☐ ☐ ☐ ☐ ☐
11. Have you felt downhearted and blue? ☐ ☐ ☐ ☐ ☐ ☐

12. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)?

☐ All of the Time ☐ Most of the Time ☐ Some of the Time ☐ A little of the Time ☐ None of the Time

9 References

1. Bloch S. Understanding troubled minds: a guide to mental illness and its treatment. Melbourne: Melbourne University Press; 2011.
2. Wessely S. Risk, psychiatry and the military. *Br J Psychiatry*. 2005;186:459-66.
3. Rundell JR, Ursano RJ. Psychiatric responses to war trauma. In: Ursano RJ, Norwood AE, editors. *Emotional aftermath of the Persian Gulf War : veterans, families, communities and nations*. Washington, D.C.: American Psychiatric Press; 1996. p. 43-81.
4. Ikin JF, Sim MR, Creamer MC, Forbes AB, McKenzie DP, Kelsall HL, et al. War-related psychological stressors and risk of psychological disorders in Australian veterans of the 1991 Gulf War. *Br J Psychiatry*. 2004;185:116-26.
5. Hoge CW, Auchterlonie JL, Milliken CS. Mental health problems, use of mental health services, and attrition from military service after returning from deployment to Iraq or Afghanistan. *Journal of the American Medical Association*. 2006;295(9):1023-32.
6. Ikin JF, Sim MR, McKenzie DP, Horsley KWA, Wilson EJ, Moore MR, et al. Anxiety, PTSD and depression in Korean War veterans fifty years after the war. *Br J Psychiatry*. 2007;190:475-83.
7. Australian Institute of Health and Welfare. The burden of disease and injury in Australia. Canberra: Australian Institute of Health and Welfare; 2007.
8. World Health Organization. The global burden of disease: 2004 update. Geneva: World Health Organization; 2008.
9. Stern SL, Dhanda R, Hazuda HP. Helplessness predicts the development of hypertension in older Mexican and European Americans. *J Psychosom Res*. 2009;67(4):333-7.
10. Kelsall HL, McKenzie DP, Sim MR, Leder K, Forbes AB, Dwyer T. Physical, psychological, and functional comorbidities of multisymptom illness in Australian male veterans of the 1991 Gulf War. *Am J Epidemiol*. 2009;170(8):1048-56.
11. McKenzie DP, Creamer M, Kelsall HL, Forbes AB, Ikin JF, Sim MR, et al. Temporal relationships between Gulf War deployment and subsequent psychological disorders in Royal Australian Navy Gulf War veterans. *Soc Psychiatry Psychiatr Epidemiol*. 2010;45(9):843-52.
12. McKenzie DP, Ikin JF, McFarlane AC, Creamer M, Forbes AB, Kelsall HL, et al. Psychological health of Australian veterans of the 1991 Gulf War : An assessment using the SF-12, GHQ-12 and PCL-S. *Psychol Med*. 2004;34(8):1419-30.

13. Zvin K, McCarthy JF, McCammon RJ, Valenstein M, Post EP, Welsh DE, et al. Health-related quality of life and utilities among patients with depression in the Department of Veterans Affairs. *Psychiatr Serv.* 2008;59:1331-4.
14. Forman-Hoffman VL, Carney CP, Sampson TR, Peloso PM, Woolson RF, Black DW. Mental health comorbidity patterns and impact on quality of life among veterans serving during the first Gulf War. *Qual Life Res.* 2005;14:2303-14.
15. Baas KD, Wittkamp KA, van Weert HC, Lucassen P, Huyser J, van den Hoogen H, et al. Screening for depression in high-risk groups: prospective cohort study in general practice. *Br J Psychiatry.* 2009;194(5):399-403.
16. Glover TA, Alberg CA. Considerations for evaluating universal screening instruments. *Journal of School Psychology.* 2007;45:117-35.
17. Clarke DM, McKenzie DP. A caution on the use of cut-points applied to screening instruments or diagnostic criteria. *J Psychiatr Res.* 1994; 28(2):185-8.
18. McKenzie DP, Clarke DM, Forbes AB, Sim MR. Pessimism, worthlessness, anhedonia and thoughts of death identify DSM-IV major depression in hospitalized medically ill. *Psychosomatics.* 2010;51:302-11.
19. Zimmerman M, Galione JN, Chelminski I, McGlinchey JB, Young D, Dalrymple K, et al. A simpler definition of major depressive disorder. *Psychol Med.* 2010;40(3):451-7.
20. Britton PC, Bossarte RM, Lu N, He H, Currier GW, Crilly J, et al. Prevalence, correlates, and symptom profiles of depression among men with a history of military service. *Soc Psychiatry Psychiatr Epidemiol.* 2011;46:607-14.
21. Izal M, Montorio I, Nuevo R, Perez-Rojo G, Cabrera I. Optimising the diagnostic performance of the Geriatric Depression Scale. *Psychiatry Res.* 2010;178(1):142-6.
22. Clarke DM, McKenzie DP, Marshall RJ, Smith GC. The construction of a brief case-finding instrument for depression in the physically ill. *Integrative Psychiatry.* 1996;10:117-23.
23. McKenzie DP, Toumbourou JW, Forbes AB, Mackinnon AJ, McMaster. B.J., Catalano RF, et al. Predicting future depression in adolescents using the Short Mood and Feelings Questionnaire: a two-nation study. *J Affect Disord.* 2011;134:151-9.
24. Goldberg DP. The detection of psychiatric illness by questionnaire. London: Oxford University Press; 1972.
25. Rhew IC, Simpson K, Tracy M, Lymp J, McCauley E, Tsuang D, et al. Criterion validity of the Short Mood and Feelings Questionnaire and one- and two-item depression screens in young adolescents. *Child Adolesc Psychiatry Ment Health.* 2010;4:8.
26. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-IV). 4th ed. Arlington, VA: American Psychiatric Association; 1994.

27. Hotopf M, David A, Hull L, Nikalaou V, Unwin C, Wessely S. Risk factors for continued illness among Gulf War veterans: a cohort study. *Psychol Med.* 2004;34:1-8.
28. Makowska Z, Merecz D, Moscicka A, Kolasa W. The validity of general health questionnaires, GHQ-12 and GHQ-28, in mental health studies of working people. *International Journal of Occupational and Medical Environmental Health.* 2002;15(4):353-62.
29. Puustinen PJ. Screening for psychological distress. In: Ohayashi H, Yamada S, editors. *Psychological distress: symptoms, causes and coping.* New York: Nova; 2012. p. 49-66.
30. Clarke DM, Smith GC, Herrman HE. A comparative study of screening instruments for mental disorders in general hospital patients. *Int J Psychiatry Med.* 1993;23:323-37.
31. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-5).* 5th ed. Arlington, VA: American Psychiatric Association; 2013.
32. World Health Organization Collaborating Centre for Mental Health and Substance Abuse. *Composite International Diagnostic Interview: CIDI-Auto 2.1 - Administrator's guide and reference.* Sydney: World Health Organization Collaborating Centre for Mental Health and Substance Abuse; 1997.
33. Clarke DM, Mackinnon AJ, Smith GC, McKenzie DP, Herrman HE. Dimensions of psychopathology in the medically ill : a latent trait analysis. *Psychosomatics.* 2000;41:418-25.
34. Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A, et al. The chronic fatigue syndrome: A comprehensive approach to its definition and study. *Ann Intern Med.* 1994;121:953-9.
35. Kelsall H, Sim M, McKenzie D, Forbes A, Leder K, Glass D, et al. Medically evaluated psychological and physical health of Australian Gulf War veterans with chronic fatigue. *J Psychosom Res.* 2006;60:575-84.
36. Fukuda K, Nisenbaum R, Stewart G, Thompson WW, Robin L, Washko RM, et al. Chronic multisymptom illness affecting Air Force veterans of the Gulf War. *JAMA.* 1998;280(11):981-8.
37. Blanchard MS, Eisen SA, Alpern R, Karlinsky JB, Toomey R, Reda DJ, et al. Chronic multisymptom illness complex in Gulf War I veterans 10 years later. *Am J Epidemiol.* 2006;163(1):66-75.
38. Unwin C, Blatchley N, Coker W, Ferry S, Hotopf M, Hull L, et al. Health of UK servicemen who served in Persian Gulf War. *Lancet.* 1999;353(9148):169-78.
39. Kelsall HL, Sim MR, Forbes AB, Glass DC, McKenzie DP, Ikin JF, et al. Symptoms and medical conditions in Australian veterans of the 1991 Gulf War: relation to immunisations and other Gulf War exposures. *Occup Environ Med.* 2004;61(12):1006-13.

40. Kelsall HL, McKenzie DP, Forbes AB, Roberts MH, Urquhart DM, Sim MR. Pain-related musculoskeletal disorders, psychological comorbidity and the relationship with physical and mental wellbeing in Gulf War veterans. *Pain*. 2013 submitted.
41. Abouzeid M, Kelsall HL, Forbes AB, Sim MR, Creamer MC. Posttraumatic stress disorder and hypertension in Australian veterans of the 1991 Gulf War. *J Psychosom Res*. 2012;72:33-8.
42. Ware JE, Kosinski MA, Keller SD. SF-12: How to score the SF-12 physical and mental health summary scales. 3rd ed. Lincoln, RI: QualityMetric Incorporated; 1998.
43. Witten IH, Frank E, Hall MA. Data mining: practical machine learning tools and techniques. 3rd ed. San Francisco, CA: Morgan Kaufmann; 2011.
44. Efroymson MA. Multiple regression analysis. In: Ralston A, Wilf HS, editors. *Mathematical methods for digital computers*. New York: Wiley; 1960. p. 191-203.
45. Everitt BE. *A whistle-stop tour of statistics*. Boca Raton, FL: Chapman & Hall / CRC; 2012.
46. Green JG, Gruber MJ, Sampson NA, Zaslavsky AM, Kessler RC. Improving the K6 short scale to predict serious emotional disturbance in adolescents in the USA. *International Journal of Methods in Psychiatric Research*. 2010;19 Suppl 1:23-35.
47. Breiman L, Friedman J, Olshen RA, Stone CJ. *Classification and regression trees*. Belmont, CA: Wadsworth; 1984.
48. Batterham PJ, Christensen H, Mackinnon AJ. Modifiable risk factors predicting major depression disorder at four year follow-up: a decision-tree approach. *BMC Psychiatry*. 2009;9:75.
49. Marshall RJ. The use of classification and regression trees in clinical epidemiology. *J Clin Epidemiol*. 2001;54:603-9.
50. Marshall RJ. A program to implement a search method for identification of clinical subgroups. *Stat Med*. 1995;14:2645-59.
51. Donath S. The validity of the 12-item General Health Questionnaire in Australia: a comparison between three scoring methods. *Aust N Z J Psychiatry*. 2001;35(2):231-5.
52. Kraemer HC. *Evaluating medical tests : objective and quantitative guidelines*. Newbury Park, CA: Sage; 1992.
53. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837-45.
54. Hosmer DW, Lemeshow S. *Applied logistic regression*. 3rd ed. New York: Wiley; 2013.

55. Koenker R. Quantile regression. Thousand Oaks, CA: Sage; 2005.
56. World Health Organization. Obesity: Preventing and managing the global epidemic. Geneva: World Health Organization; 2000.
57. Feinstein AR. Multivariable analysis: an introduction. New Haven, CT: Yale University Press; 1996.
58. Simon GE, Von Korff M. Medical co-morbidity and validity of DSM-IV depression criteria. *Psychol Med*. 2006;36(1):27-36.
59. Swinkels CM, Ulmer CS, Beckham JC, Buse N, Calhoun PS. The association of sleep duration, mental health, and health risk behaviors among U.S. Afghanistan/Iraq era veterans. *Sleep*. 2013;36(7):1019-25.
60. Calohan J, Peterson K, Peskind ER, Raskind MA. Prazosin treatment of trauma nightmares and sleep disturbance in soldiers deployed in Iraq. *J Trauma Stress*. 2010;23(5):645-8.
61. Pearlin LI, Scholler C. The structure of coping. *J Health Soc Behav*. 1978;19:2-21.
62. Vinokur AD, Pierce PF, Lewandowski-Romps L, Hobfoll SE, Galea S. Effects of war exposure on air force personnel's mental health, job burnout and other organizational related outcomes. *J Occup Health Psychol*. 2011;16(1):3-17.
63. Pietsch K, Allgaier A-K, Fruhe B, Rohde S, Hosie S, Heinrich N, et al. Screening for depression in adolescent paediatric patients: validity of the new Depression Screener for Teenagers (DesTeen). *J Affect Disord*. 2011;134:266-71.
64. Delgadillo J, Payne S, Gilbody S, Godfrey C, Gore S, Jessop D, et al. How reliable is depression screening in alcohol and drug users? A validation of brief and ultra-brief questionnaires. *J Affect Disord*. 2011;133:69-75.
65. McKenzie DP, Clarke DM, Low LH. A method of constructing parsimonious diagnostic and screening tests. *International Journal of Methods in Psychiatric Research*. 1992;2:71-9.
66. Jonas BS, Franks P, Ingram DD. Are symptoms of anxiety and depression risk factors for hypertension? longitudinal evidence from the National Health and Nutrition Examination Survey I Epidemiologic Follow-up Study. *Arch Fam Med*. 1997;6:43-9.
67. De Hert M, Correll CU, Bobes J, Cetkovich-Bakmas M, Cohen D, Asai I, et al. Physical illness in patients with severe mental disorders. I. prevalence, impact of medications and disparities in health care. *World Psychiatry*. 2011;10:52-77.
68. Steptoe A. Depression and physical illness. Cambridge, UK: Cambridge University Press; 2007.

69. Mercer SW, Gunn H, Bower P, Wyke S, Guthrie B. Managing patients with mental and physical multimorbidity: changes are needed in policy, research and practise [editorial]. *Br Med J*. 2012;345:7.
70. Chen JJ, Marsh LC. Depression in Parkinson's disease: identification and management. *Pharmacotherapy*. 2013;33(9):972-83.
71. Donker T, van Straten A, Marks I, Cuijpers P. Brief self-rated screening for depression on the Internet. *J Affect Disord*. 2010;122(3):253-9.
72. Christensen H, Griffiths K. *The MoodGYM: overcoming depression with CBT and other effective therapies*. London: Vermillion; 2011.
73. Rona RJ, Jones M, French C, Hooper R, Wessely S. Screening for physical and psychological illness in the British Armed Forces: I: The acceptability of the programme. *J Med Screen*. 2004;11(3):148-52.
74. Cully JA, Graham DP, Kramer JR. A 2-item screen for depression in rehabilitation inpatients. *Archives of Physical Medical Rehabilitation*. 2005;86(3):469-72.
75. Clarke DM, Currie KC. Depression, anxiety and their relationship with chronic diseases: a review of the epidemiology, risk and treatment evidence. *Med J Aust*. 2009;190:S54-S60.
76. Kroenke K, Bair MJ, Damush TM, Wu J, Hoke S, Sutherland J, et al. Optimized antidepressant therapy and pain self-management in primary care patients with depression and musculoskeletal pain: a randomized controlled trial. *Journal of the American Medical Association*. 2009;301:2099-110.
77. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-13.
78. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol Med*. 2002;32:959-76.
79. Furukawa TA, Kessler RC, Slade T, Andrews G. 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*. 2003;33:357-62.