



Research report

Screening and case finding for depression in offender populations: A systematic review of diagnostic properties

Catherine E. Hewitt^{a,*}, Amanda E. Perry^b, Barbara Adams^c, Simon M. Gilbody^d

^a Department of Health Sciences, University of York, York YO10 5DD, UK

^b Centre for Criminal Justice Economics & Psychology, University of York, York YO10 5DD, UK

^c Department of Health Sciences, University of York, York YO10 5DD, UK

^d Department of Health Sciences and Hull York Medical School (HYMS), University of York, York YO10 5DD, UK

ARTICLE INFO

Article history:

Received 1 March 2010

Received in revised form 22 June 2010

Accepted 22 June 2010

Available online 23 July 2010

Keywords:

Depression

Sensitivity

Specificity

Diagnostic accuracy review

Offender

Prison

ABSTRACT

Background: Diagnosis of depression in offender populations is particularly difficult for health professions because of the many vulnerable complex problems associated with this population. As offender populations represent an 'at risk population', one feasible approach is the use of brief standardised mood assessments that can be either self-completed or completed by a non-specialist.

Aims: To review the diagnostic accuracy of brief psychometric instruments to identify depression in offender populations.

Method: The authors searched five electronic databases from inception to March 2009 and examined reference lists to identify the relevant literature. The authors included studies comparing the accuracy of any brief psychometric instrument to identify depression in offender populations with a standardised diagnostic interview conducted according to internationally recognised criteria. Two reviewers independently reviewed each article to assess inclusion, extract relevant study characteristics and data.

Results: In total, thirteen studies met the inclusion criteria. Instruments validated in offender populations included both general depression questionnaires as well as specific measures that had been developed for use in offender populations. The most frequently validated instruments were the General Health Questionnaire (GHQ) and the Referral Decision Scale (RDS).

Conclusions: A number of different tools were identified in the review which could perhaps serve as a benchmark for the identification of depression in offender populations.

© 2010 Elsevier B.V. All rights reserved.

1. Background

Depression accounts for the greatest burden of disease among all mental health problems, and is expected to become the second highest amongst all general health problems by 2020 (Murray and Lopez, 1996). Depression is at least as common in offender populations as in the general population, and is often missed or misdiagnosed (Brooke et al., 1996). Under-recognised and poorly managed mood disorders con-

tribute to the poor general health of offenders (Cooper and Berwick, 2001). Rates of self-harm and completed suicide are a major public health problem in the penal system and amongst offender populations living within the community (Shaw et al., 2004). The presence of depression acts as a risk factor for self-harm and suicide and is amenable to treatment with evidence-supported interventions (Shaw et al., 2004).

Access to professionals skilled in psychological assessment and the diagnosis of depression is limited for offender populations. As offender populations represent an 'at risk population', one feasible approach is the use of brief standardised mood assessments that can be either self-completed or completed by a non-specialist. A wide range of instruments

* Corresponding author. Tel.: +44 1904 321374; fax: +44 1904 321920.
E-mail address: ceh121@york.ac.uk (C.E. Hewitt).

exist and their use is advocated in UK primary care under the General Medical Services (GMS) Quality and Outcomes Framework (QOF) (BMA and NHS Employers, 2006). Instruments such as the Patient Health Questionnaire nine-items (PHQ9) have passed into common use in UK Primary Care (Kroenke et al., 2001; Vedavanam et al., 2009). The diagnostic properties of brief instruments are broadly acceptable in the general population, but cannot be assumed in offender populations (Mitchell and Coyne, 2007; Gilbody et al., 2007; Williams et al., 2002). There are several reasons why diagnostic properties (such as sensitivity and specificity) might not directly translate from general to offender populations, including problems with the over-diagnosis of depression when terms which have different nuances for offenders (such as 'guilt') could be given undue weight in standardised instruments. Respondent biases and differing clinical presentations make it essential that specific validation studies are sought in offender populations prior to their implementation.

Systematic reviews addressing the fundamental diagnostic epidemiology of brief psychometric instruments have been prepared in relation to depression in general and in relation to self-harm in offender populations (Mitchell and Coyne, 2007; McMillan et al., 2007). However, there have been (to our knowledge) no systematic reviews of the properties of brief standardised depression instruments for depression in offending populations. Against this background and in the absence of an existing review, we comprehensively and systematically synthesised the evidence. The aims of the review were to estimate the diagnostic accuracy of each psychometric instrument and to compare the accuracy between instruments.

2. Methods

We used state of the art methods in the conduct of systematic reviews using guidance laid down by the Centre for Reviews and Dissemination 4th report (third edition), with specific adaptations in relation to reviews of diagnostic performance (Centre for Reviews and Dissemination, 2009; Deville et al., 2002).

3. Data sources and searches

Searches were undertaken across five electronic databases in criminal justice, psychology and health. The searches were structured to capture three concepts: offender populations, depression, and screening or diagnosis or identification tools. To ensure the search was as sensitive as possible in retrieving relevant records, a combination of subject headings (thesaurus terms) and text words were used in the strategy. Sensitivity was also enhanced by searching using a number of named screening and diagnosis tools and instruments. Precision (focus) was enhanced by searching for text words about screening for depression in close proximity. Letters, editorials and notes were excluded from the results, where possible. This approach was used to achieve the required focus on the search results on research studies. All databases were searched from their inception until March 2009. No language or other restrictions were applied. An example of our search strategy is provided in at the end of the paper. Reference lists of all studies were inspected to ensure that all potentially relevant studies had been identified.

4. Study selection and inclusion criteria

Records were downloaded from the databases and were loaded into an endnote bibliographic software database. As records were loaded they were automatically de-duplicated. Two of the authors screened titles and abstracts to identify potentially eligible studies. Any disagreements were resolved by consensus or deferred to a third party, if necessary. Full papers for potentially eligible studies were obtained and assessed for inclusion independently by two of the authors. Articles were included if they prospectively compared the performance of any brief psychometric instrument to identify depression in offender populations with a standardised diagnostic interview conducted according to internationally recognised criteria (e.g. International Classification of Diseases (ICD) system). We included studies which used a brief instrument which were either self-report or administered by a lay person without specific/in-depth training. Instruments had to either focus exclusively on the presence or absence of depressive symptoms and syndromes which examine depression and other mood symptoms (such as anxiety). We used an operational working definition of depression as a syndrome constituting 'sad, despairing mood; decrease of mental productivity and reduction of drive; and retardation or agitation in motor behavior' (Lorr et al., 1967).

5. Data extraction and quality assessment

All data were extracted independently by two reviewers. We assessed the quality of studies according to accepted criteria using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) instrument (Whiting et al., 2004). Important plausible sources of bias include the use of a representative spectrum sample of patients, the application of diagnostic gold standard interviews, and blinding of gold standard raters to scores on the case identification test. We used our own refinements of the QUADAS instrument for use in reviews of diagnostic instruments (Mann et al., 2008).

6. Data synthesis and analysis

For each instrument, the range in sensitivity and specificity was calculated. There were insufficient data and substantial heterogeneity (in terms of different instruments and populations) such that a quantitative synthesis (meta-analysis) was not performed. We present a narrative overview of key design elements (population, setting, instrument, diagnostic standard and methodological quality). In general, when any test is used there are four possible outcomes:

- When a person has the condition the test may be positive (true positive);
- When a person has the condition the test may be negative (false negative);
- When a person does not have the condition the test may be negative (true negative);
- When a person does not have the condition the test may be positive (false positive).

Using this information we constructed a series of 2 by 2 tables of results of the case finding instrument versus results according to a diagnostic gold standard to summarise the data

Table 1

Summary 2 by 2 table.

Case finding	Gold standard		
	+	+	–
	+	True positive	False positive
	–	False negative	True negative

(Table 1). We then plotted this information on a graph to allow a visual comparison of the instruments identified. Sensitivity and specificity are dependent upon one another, if one value decreases the other value increases. Hence increasing the cut point used increases or decreases the sensitivity and specificity of the test. In some studies multiple cut points were reported for each instrument and hence we needed to decide which cut point should be selected. Youden's index is one way to attempt to summarise test performance into a single numeric value to aid decision making regarding cut points (Youden, 1950). We decided to compare the instruments in three ways by selecting the cut point that maximised sensitivity, maximised specificity and the cut point with the highest Youden's index value.

7. Results

In total, 1396 potentially relevant studies were identified from the searches, of which 58 were selected for full assessment (Fig. 1). 13 studies (5565 individuals) met the inclusion criteria and were included in the review.

8. Characteristics of included studies

Studies were published between 1989 and 2008 and were undertaken in a variety of countries (Table 2): 8 in the US, 2 in the UK, 1 in Australia, Canada and Denmark.

Table 2

Summary of the number of studies using each instrument.

Instrument	No. of studies	Versions used
BJMHS	2	BJMHS (split by gender) BJMHS-R 8 and 12 item versions (split by gender)
CODSI	3	6 item for any mental disorder (CODSI-MD; cut point: 3; split by gender); 3 item for severe mental disorder (CODSI-SMD; cut point: 2; split by gender)
GHQ	3	12 item (cut point: 2; 2 time points); 28 item (cut points: 2 to 17); 30 item GHQ-30 (cut points: 5, 11, 16); GHQ-short (includes items 3, 16, 21, 24 of the GHQ-30; cut points: 1 to 3)
GSS	2	Cut point: 2 (split by gender) for MD Cut point: 5 (split by gender) for SMD
JSAT	1	Single cut point
MAYSI-2	1	Cut points: 3 to 6
MDSIS	1	Cut points: 1 to 3
MFQ	1	Long version (MFQ): Cut points: 27, 29, 30; Short version (SMFQ): 8, 10, 11
MHSF	2	Cut points: 2 to 6, and 11 for SMD Cut points: 3 and 11 (split by gender)
PISP	1	Single cut point
RDS	3	Cut points: 1 to 4; Unclear; Validation and development samples
RDS or PISP	1	Single cut point

Within the diagnostic accuracy studies, different reference standards were used: Nine (69%) of the studies used Diagnostic and Statistical Manual of Mental Disorders (DSM) classifications, 3 (33%) used International Statistical Classification of Diseases

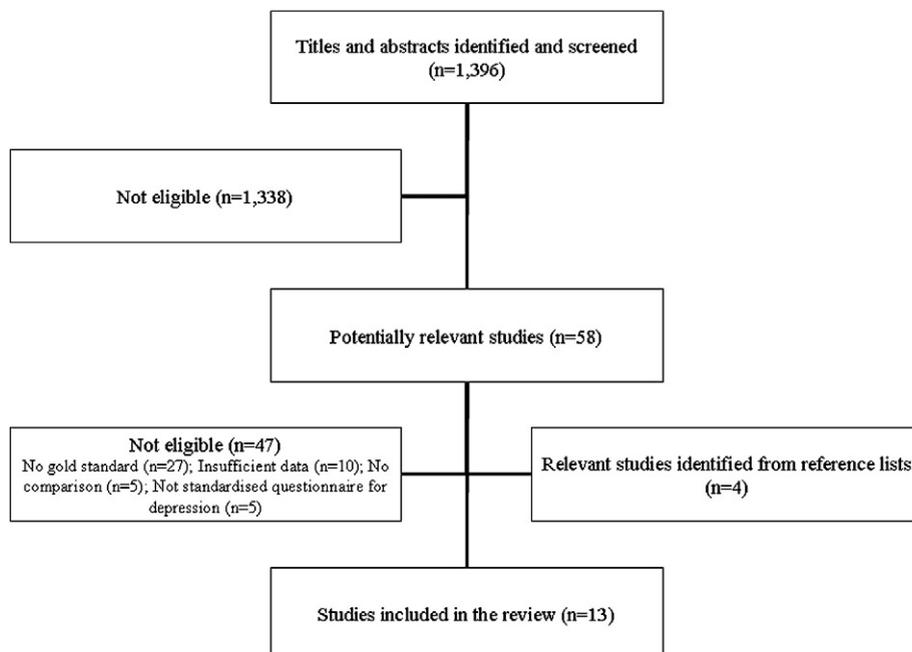
**Fig. 1.** Flow of studies.

Table 3
Summary of brief psychometric instruments.

Instrument	Sample	Type of instrument	No. items	Score range	Time frame, completion time
BJMHS revision of the RDS	Adults; authors recommend for men only	Specific; officer administered with little or no training required	8 (yes or no questions)	No scoring required. Referred if answer 'yes' to item 7 or 'yes' to item 8 or 'yes' to at least 2 items 1 to 6	Currently 3 min
CODSI (CODSI-MD or CODSI-SMD)	Adults	Specific;	6 for mental disorder 3 for severe mental disorder	Unclear	At some point in their life Unclear
GHQ (GHQ-12 or GHQ-28 or GHQ-30)	Adults	Generic; self-complete	30; 12	0 to 90 (GHQ-30) 0 to 84 (GHQ-28) 0 to 36 (GHQ-12)	Past few weeks 2 to 10 min
GSS	Adolescents and adults	Specific; self-or staff administration	23; 4 questions with 5 sub-questions each with 4 point scale response ("never", "1+ years ago", "2 to 12 months ago" and "past month"); 3 further questions	0 to 20	At some point in their life (responses are given in terms recency of the problem described) 5 min
JSAT	Adults	Specific; screeners should have graduate training in psychopathology and assessment + specific JSAT training	8 sections: demographics, legal situation, violence issues, social background, substance use, mental health treatment, suicide and self-harm issues, and mental health status	No scoring and thus no summary scores and no score-based decision rules	Unclear 10 to 20 min
MAYSI-2	Youths 12 to 17 who may have special mental health needs	Specific; self-complete	52 in total; 9 depression/anxiety items	0 to 9 for depression/anxiety subscale	Past few months 10 to 15 min
MDSIS	Adults	Specific; officer administered	12 interview based questions and 7 observational items	No scoring required. A single inappropriate response indicates further evaluation.	At some point in their life 5 min
MFQ (including SMFQ)	Children and adolescents 8 to 18 to detect major depression SMFQ: Children and adolescents 6 to 17	Specific; self-complete Child and parental version	34 (MFQ) 13 (SMFQ) Responses given on 3-point scale ("not true", "sometimes true" and "true")	0 to 68 (MFQ) 0 to 26 (SMFQ)	Preceding 2 weeks 5 to 10 min (MFQ) 5 min (SMFQ)
MHSF	Adults	Specific; self-or staff administration	18 (yes or no questions)	0 to 18 A qualified mental health specialist should be consulted about any "yes" response to questions 3 to 17.	At some point in their life 5 min
PISP	Adults	Specific; staff	31 in total; 3 that address mental health	Yes/no responses used to determine the necessity of medical or mental health intervention	At some point in their life Unclear
RDS	Adults Designed to detect schizophrenia, manic-depressive illness and major depression	Specific; officer administered with little or no training required	15 (3 sub-sections with 5 yes or no questions in each)	0 to 5 for each sub-scale	At some point in their life 10 min

and Related Health Problems (ICD) and 1 (11%) used Research Diagnostic Criteria (RDC).

Instruments validated in offender populations included both general depression questionnaires as well as specific measures that had been developed for use in offender populations. These

are shown in Table 2 and comprise: General Health Questionnaire (GHQ), Brief Jail Mental Health Screen (BJMHS), Referral Decision Scale (RDS), Mental Disability/Suicide Intake Screen (MDSIS), Mood and Feelings Questionnaire (MFQ) – short and long forms, Massachusetts Youth Screening Instrument

Table 4
Characteristics of included studies.

Author(s), year, country, instrument	Study sample, age, gender	Administration	Interview type, gold standard	Type of classification	Sample size, baseline prevalence
(Steadman et al., 2007), US, Brief Jail Mental Health Screen – Revised (BJMHS-R)	Detainees admitted to one of four county prisons. Adults Males: 206 (44%)	Questionnaire administered within 24 h of initial booking	Structured Clinical Interview for DSM (SCID)	Major depressive disorder; depressive disorder NOS; bipolar disorder (I and II and NOS); schizoaffective disorder; schizophreniform disorder; brief psychotic disorder; delusional disorder; and psychotic disorder NOS	Males n = 206 33 (16%) Females n = 258 63 (20%)
(Harrison and Rogers, 2007), US, Referral Decision Scale (RDS); Mental Disability/Suicide Intake Screen (MDSIS)	Participants recruited from single prison through posters. Participants offered \$5 for participation. Adults Males: 49 (49%)	Had to be detained for less than 4 weeks. Interviewed in private room.	Schedule of Affective Disorders and Schizophrenia (SADS), DSM-IV	Depression	Total n = 100 13 (13%)
(Kuo et al., 2005), US, Mood and Feelings Questionnaire (MFQ); MFQ-Short (SMFQ); Massachusetts Youth Screening Instrument (MAYSI-2)	Participants recruited from single prison. Youths: 13–17 years Males: 170 (75%)	Had been detained at least 8 h. Interviewed in a private room.	Voice-Diagnostic Interview Schedule for Children (V-DISC), DSM-IV	Major Depressive Disorder (MDD)	Total n = 228 32 (14%) 50 given V-DISC – quasi random sample
(Hurley and Dunne, 1991), Australia, General Health Questionnaire (GHQ-12)	Participants recruited from single women's prison Adults: 17–55 years Males: 0 (0%)	Subjects were assessed in an interview.	Structured Clinical Interview for DSM-III-R (SCID)	Psychiatric disorder	Time 1 n = 92 49 (53%) Time 2 n = 49 25 (51%)
(McLearen and Ryba, 2003), US, Prisoner Intake Screening Procedure (PISP); Referral Decision Scale (RDS)	Participants were selected from all new admissions to a single prison Adults: 16–60 years Males: 95 (100%)	Booking officers administered the inventory to all inmates upon entrance to the facility.	Schedule of Affective Disorders and Schizophrenia Change version (SADS-C), RDC	Severely mentally ill (including major depression, bipolar disorder and schizophrenia)	Total n = 95 11 (12%)
(Shaw et al., 2003), UK, General Health Questionnaire (GHQ)	Attendees at Manchester and Preston magistrates court Adults Males: 1123 (86%)	Subjects were assessed in an interview.	Schedule for Clinical Assessment in Neuropsychiatry (SCAN), ICD	Depression	Total n = 1306 68 (5%) Interviewed subjects with GHQ ≥ 4 or PSQ ≥ 1 and a random sample of 16% GHQ/PSQ screen negatives

(Andersen, 2004), Denmark, General Health Questionnaire (GHQ-28)	Participants were chosen at random from lists of all new prisoners using a randomisation list Adults: over 18 Males: 90%	Most subjects were interviewed the day after imprisonment and by day six all subjects had been examined.	Present State Examination (PSE), ICD-10	At least one ICD-10 disorder (sections 20, 30 and 40)	Total n = 184 75 (41%)
(Sacks et al., 2007a), US, Mental Health Screening Form (MHSF); Global Appraisal of Individual Needs Short Screener (GSS); Co-occurring Disorders Screening Instrument for Severe Mental Disorders (CODSI-SMD) Pilot sample	Consecutive new admissions to prison substance abuse treatment programs across participating CJDATS research centres Adults: 16 to 68 Males: 75 (75%)	Instruments administered in 2 face-to-face sessions conducted within 1 month of each other.	Structured Clinical Interview for DSM-IV (SCID)	Mental disorders Severe mental disorders (including major depression, schizophrenia and bipolar disorder)	Total n = 100
(Sacks et al., 2007b), US, Mental Health Screening Form (MHSF); Global Appraisal of Individual Needs Short Screener (GSS); Co-occurring Disorders Screening Instrument for Severe Mental Disorders (CODSI-SMD) Validation sample	Consecutive new admissions to prison substance abuse treatment programs across participating CJDATS research centres Adults Males: 106 (59%)	Instruments administered in 2 face-to-face sessions conducted within 1 month of each other.	Structured Clinical Interview for DSM-IV (SCID)	Mental disorders Severe mental disorders (including major depression, schizophrenia and bipolar disorder)	<i>Any mental disorder</i> Total n = 180 141 (78%) <i>Severe mental disorders</i> Total n = 180 75 (42%)
(Duncan et al., 2008), US, Co-occurring Disorders Screening Instruments – any mental disorder (CODSI-MD); Co-occurring Disorders Screening Instruments – severe mental disorder (CODSI-SMD)	New admissions to prison substance abuse treatment programs Adults Males: 207 (59%)	Within one month of the intake interview and screening battery participants were interviewed.	Structured Clinical Interview for DSM-IV (SCID)	Any mental disorder and severe mental disorder	<i>Any mental disorder</i> Total n = 353 253 (72%) African American n = 96, 67 (70%) Latino n = 120, 88 (74%) White n = 137, 98 (72%) <i>Severe mental disorder</i> Total n = 353 110 (31%) African American n = 96, 25 (26%) Latino n = 120, 35 (29%) White n = 137, 50 (37%) Major depression n = 258 36 (14%)
(Teplin and Swartz, 1989), US, Referral Decision Scale (RDS)	Participants from a larger prevalence study. A stratified random sample of male detainees who entered a single prison over a year. Adults Males: 728 (100%)	Subjects were interviewed in a soundproof, private glass booth in the intake area	National Institute for Mental Health Diagnostic Interview Schedule (NIMH-DIS), DSM-III	Schizophrenia, manic depressive disorder and major depression	

(continued on next page)

Table 4 (continued)

Author(s), year, country, instrument	Study sample, age, gender	Administration	Interview type, gold standard	Type of classification	Sample size, baseline prevalence
Validation sample	Sentenced inmates in a single prison Adults: three quarters were age 30 or younger	Unclear	National Institute for Mental Health Diagnostic Interview Schedule (NIMH-DIS), DSM-III	Schizophrenia, manic depressive disorder and major depression	Major depression n = 1,149 56 (5%)
(Steadman et al., 2005), US, Brief Jail Mental Health Screen (BJMHS)	Males: unclear Participants were jail detainees admitted to one of four county jails Adults Males: 211 (59%)	Subjects completed the questionnaires on admission to the jail. Interviews were undertaken within 96 h of admission.	Structured Clinical Interview for DSM-IV (SCID)	Serious mental illness (major depressive disorder, depressive disorder not otherwise specified, bipolar disorder (I, II, and not otherwise specified), schizophrenia disorder, schizoaffective disorder, schizophreniform disorder, brief psychotic disorder, delusional disorder, and psychotic disorder not otherwise specified)	Males n = 211 58(27%) Females n = 146 61(42%)
(Nicholls et al., 2004), Canada, Jail Screening Assessment Tool (JSAT)	Participants from another study were sampled. The sample was selected to ensure that an adequate number of inmates with mental health problems were sampled. Subjects were from a single prison. Adults Males: 0 (0%)	At the intake interview the JSAT (including the BPRS-E) was completed. Within 1 to 8 days from the intake interview, women were interviewed.	Structured Clinical Interview for DSM-IV non-patient version (SCID-I/NP)	Major DSM-IV axis I diagnoses	Total n = 29 12 (41%)

(MAYSI-2), Mental Health Screening Form (MHSF), Global appraisal of individual needs Short Screener version 1 (GSS) and Co-occurring Disorders Screening Instruments (CODSI) – any mental disorder and severe mental disorder.

Tables 3 and 4 provide two summaries detailing different aspects of the psychometric instruments and the characteristics of the included studies. From our 13 instruments, 2 were specifically designed to be used on adolescents and children under the age of 18 years (MAYSI, and the MFQ). For the other 12 instruments all were mainly used with male offenders with the exception of two studies which used females (Hurley and Dunne, 1991; Nicholls et al., 2004).

Most of the instruments could be self-administered and were completed between eight hours and within one month of admission to a secure establishment, with the number of items ranging from eight on the BJMHS to 52 on the MAYSI-2. Ranges of different cut off scores were used to define depression across the studies and the time taken to complete the instruments ranged from 3 to 15 min.

The most commonly used gold standard criteria was taken from the DSM (SCID), DSM IV or the SCAN using the ICD-10 classification system. Definitions of depression and diagnostic criteria varied across the studies ranging from the inclusion of major depressive disorders including schizophrenia, psychotic

Table 5
Summary of data.

Author, year	Instrument	Sensitivity	Specificity	Graph	
Steadman et al., 2007	BJMHS-R (12)	0.67	0.73	All	
	BJMHS-R – (8)	0.64	0.84	All	
Harrison and Rogers, 2007	MDSIS	0.46	0.92	Sensitivity; tradeoff	
		0.00	0.99	Specificity	
	RDS	1.00	0.24	Sensitivity	
		0.46	0.84	Specificity	
Kuo et al., 2005	MAYSI-2	0.85	0.49	Tradeoff	
		0.71	0.80	Sensitivity; tradeoff	
		0.29	0.96	Specificity	
	MFQ	0.71	0.92	All	
	SMFQ	1.00	0.72	Sensitivity; tradeoff	
Hurley and Dunne, 1991 McLearn and Ryba, 2003	GHQ-12	0.57	0.84	Specificity	
		0.88	0.84	All	
	PISP	0.45	0.96	All	
	RDS	0.73	0.84	All	
Shaw et al., 2003	RDS or PISP	0.91	0.83	All	
		GHQ-30	1.00	0.35	Sensitivity
			0.49	0.95	Specificity
	GHQ-short	0.80	0.90	Tradeoff	
		0.97	0.41	Sensitivity	
Andersen, 2004	GHQ-28	0.66	0.83	Specificity	
		0.90	0.67	Tradeoff	
		0.91	0.10	Sensitivity	
		0.40	0.86	Specificity	
Sacks et al., 2007a	CODSI-MD	0.59	0.76	Tradeoff	
		0.86	0.30	Sensitivity	
		0.74	0.55	Specificity; tradeoff	
	CODSI-SMD	0.56	0.77	Sensitivity	
		0.50	0.92	Specificity	
	GSS	0.51	0.91	Tradeoff	
		0.92	0.20	Sensitivity	
		0.38	0.96	Specificity; tradeoff	
MHSF	0.97	0.30	Sensitivity		
Teplin and Swartz, 1989	RDS	0.42	0.94	Specificity; tradeoff	
		0.92	0.98	Sensitivity; tradeoff	
		0.79	0.99	Specificity	
Steadman et al., 2005	BJMHS	0.66	0.77	All	
Nicholls et al., 2004	JSAT	0.75	0.71	All	
Duncan et al., 2008	CODSI-MD	0.81	0.65	Sensitivity; tradeoff	
	CODSI-MD	0.72	0.68	Specificity	
	CODSI-SMD	0.55	0.91	All	
	CODSI-MD	1.00	0.20	Sensitivity	
Sacks et al., 2007b		0.07	1.00	Specificity	
		0.84	0.80	Tradeoff	
		0.61	0.90	All	
	CODSI-SMD	0.95	0.23	Sensitivity	
		0.25	0.85	Specificity	
		0.62	0.68	Tradeoff	
		0.95	0.28	Sensitivity	
	MHSF	0.43	0.89	Specificity	
		0.75	0.68	Tradeoff	

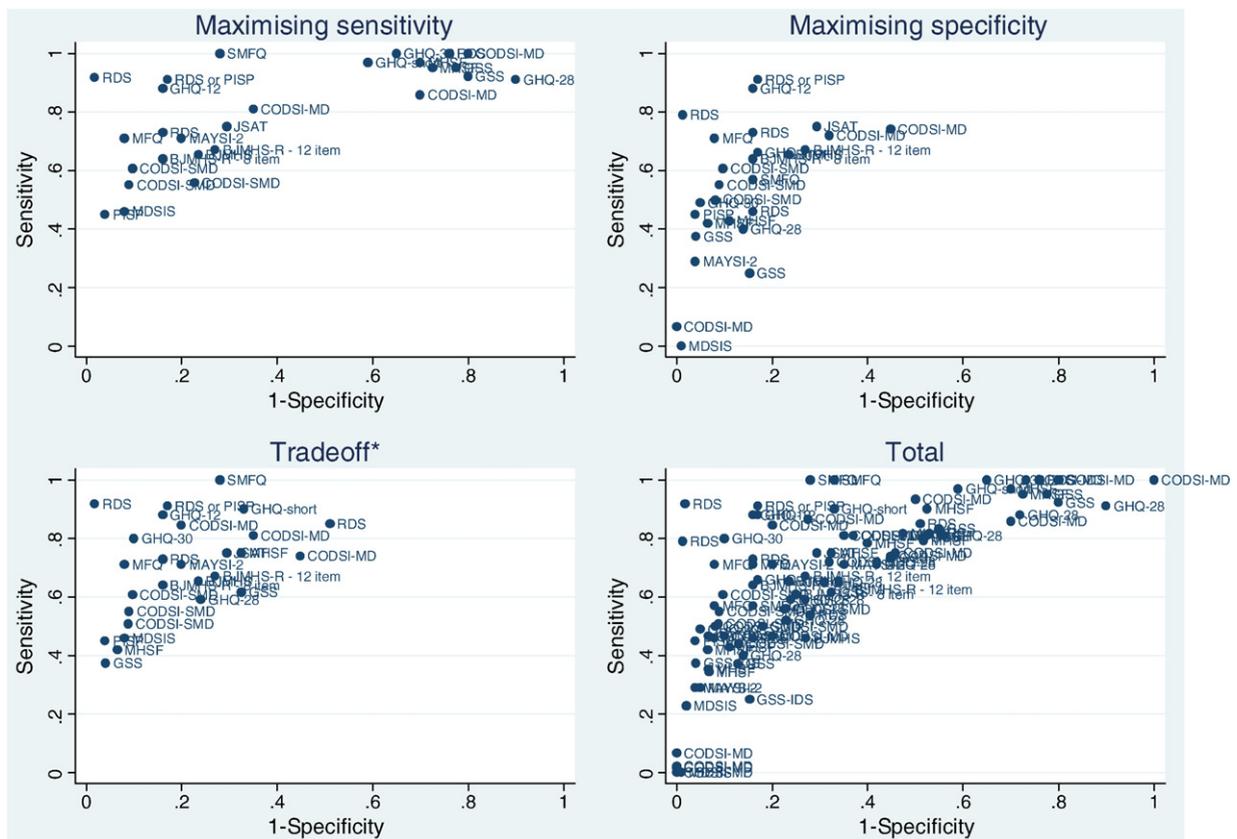


Fig. 2. Summary of sensitivity and specificity. *The tradeoff graph represents data selected on the basis of Youden's index.

and delusional disorders to having any mental disorder or severe mental disorder as classified by the SCID.

8.1. Sensitivity and specificity of the instruments

From Fig. 2 and Table 5 we can see that the instruments that appear to perform the best in terms of sensitivity and specificity are the RDS, the combined RDS and PISP and the GHQ-12. Irrespective of the choice of cut point (in terms of maximising sensitivity or specificity or by using Youden's index) the findings are consistent. We must be cautioned in over analysing the findings from the graphs in Fig. 2 as we have not accounted for the methodological quality of the studies or the populations under study and there are very few studies included.

9. Discussion

This is, to our knowledge, the first attempt to apply state of the art systematic review methods to the evaluation of screening instruments for depression in offender populations. We have highlighted the breadth and strengths of existing literature, and identified those areas where more work is needed in this neglected area of research.

Our main finding is that 13 studies have validated a case finding/screening instrument against a recognised international diagnostic gold standard. A range of different types of

instrument have been validated in this way. These instruments included both generic depression questionnaires applied in offender populations, and also instruments 'specifically designed' for use in offender populations. The most frequently used generic instrument was the GHQ (an anxiety and depression questionnaire) and the RDS (designed specifically for offenders). These were used in a range of offender populations (including youths and adults, in both remand and following sentencing), and they reported a range of different cut off scores and study characteristics in comparison to diagnostic gold standards (such as the DSM-IV SCID; Schedule for Assessment of Affective Disorders and Schizophrenia – SADS and the IC-10 SCAN).

We have also examined the properties of these instruments with respect to their ability to identify depression (sensitivity) and their ability to exclude those without depression (specificity). These are not fixed constructs, and the relative values of sensitivity and specificity will vary inversely according to the optimum cut point that is chosen. We found that instruments can be made more sensitive by choosing a low cut point, but that this occurs at the expense of reducing specificity and resulting in more false positives being identified by use of the instruments. Our study is novel in using ROC curve methods to help clinicians choose between instruments and to choose their optimum cut point. To our knowledge, this is the first time this body of literature has been presented in this way. Our main finding is that general

depression questionnaires (such as the GHQ) can produce good values of sensitivity and specificity (0.88 and 0.84 respectively) at their optimum cut point.

Our review also highlighted several gaps in the research literature. For example most studies focussed on the use of instruments in male offending populations, whilst only 2 studies examined their properties amongst female offenders. Female offenders represent a smaller but nonetheless important group. The incidence of problems such as self-harm and personality difficulties are potentially different in this population and the extrapolation of psychometric values from male offending populations should be undertaken with caution and not assumed. More research is needed in these populations to validate available depression screening measures.

A further limitation of our study was the variability in the definition of depression that was used. Although we sought to maximise the rigour of our case definition by reference to an internationally recognised gold standard, several studies used different severities of depression or conflated depression with other related psychiatric disorders such as anxiety. For this reason, it is difficult to make specific recommendations on the ability of instruments to identify more tightly-defined depression. More research is needed to delineate the specific ability of some potentially promising instruments to identify more narrowly-defined clinical depression.

From a practitioner perspective a number of issues emerge from our review which relate to the utility of these instruments. In an offender population evidence suggests that the majority of suicides and instances of self-harm occur within a one month period of admission. Instruments that identify few false positive and are quick and easy to administer are therefore imperative to direct allocated resources to those most in need (Perry and Gilbody, 2009).

The implementation of such a policy will require the use of instruments to detect psychological disorder by those with little or no training. Our review provides some indication of which instruments (such as the GHQ) might fulfil this role, by being brief and with minimal training requirements prior to their administration. For several more specialist instruments, the training requirement was rarely commented upon and to this extent it is unknown what requirements would be needed.

Notwithstanding these limitations these studies could serve as a benchmark for future health professionals who require standardised cut off scores for this particularly vulnerable population. Our main recommendation for future research in this area is that instruments should be validated against a gold standard and that the full range of false positives and false negatives should be stated in line with more general recommendations in ensuring the improvement in quality of reporting of diagnostic studies. We note that there are several relatively new brief instruments (such as the Patient Health Questionnaire 9 – PHQ9) which have now been widely validated in primary care and hospital settings (Spitzer et al., 1999; Gilbody et al., 2007). These are promising in offender populations in that they have a low training requirement and can be readily administered by a range of skilled and non-skilled prison staff. We would recommend their validation in line with the methods described in this review as a matter of some urgency.

Role funding source

Nothing declared.

Conflict of interest

All authors declare that they have no conflicts of interest.

MEDLINE – 1950 to February Week 3 2009, plus MEDLINE in process and unindexed records (searched on 02/03/09)

1. Depression/(50582);
2. depress\$.ti,ab. (228220);
3. Depressive disorder/(46975);
4. depressive disorder, major/(10057);
5. Melanchol\$.ti,ab. (1992);
6. (anxiety or anxious).ti,ab. (73382);
7. Anxiety/(37274);
8. or/1–7 (312017);
9. offender\$1.ti,ab. (5320);
10. Prisoner\$.ti,ab. (3560);
11. Criminal\$.ti,ab. (10010);
12. Prison\$1.ti,ab. (5290);
13. Prisoners/(8966);
14. Prisons/(5679);
15. Juvenile Delinquency/or Crime/(15761);
16. (secure adj2 (placement or accommodation or facilit\$ or care or unit\$ or centre\$ or center\$ or home\$)).ti,ab. (395);
17. high dependency unit\$.ti,ab. (246);
18. (prison\$ or jail\$ or gaol\$ or reformato\$ or penitentiari\$).ti,ab. (9066);
19. Reoffender\$.ti,ab. (4);
20. (forensic or criminal).jw. (16692);
21. or/9–20 (54883);
22. exp Diagnosis/(4637495);
23. ((General Health adj3 (Inventory or Questionnaire or scale or index or checklist or interview)) or ghq).ti,ab. (3081);
24. (Beck adj3 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab. (4797);
25. (BDI or bai).ti,ab. (3096);
26. ((State adj2 anxiety adj2 depression) or SAD).ti,ab. (3951);
27. (Hospital adj3 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab. (2171);
28. HADS.ti,ab. (1047);
29. (Hamilton adj3 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab. (5075);
30. HRSD.ti,ab. (334);
31. (Zung adj3 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab. (443);
32. SDS.ti,ab. (49879);
33. (Profile adj3 mood states).ti,ab. (1218);
34. POMS.ti,ab. (879);
35. (Centre adj2 Epidemiological studies adj3 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab. (45);
36. (CES-D or CESD).ti,ab. (1511);
37. (Symptom Checklist adj3 revised).ti,ab. (376);
38. SCL-90-R.ti,ab. (720);
39. (Brief symptom adj3 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab. (675);
40. BSI.ti,ab. (1097);

41. ((Inventory or Questionnaire or scale or index or checklist or interview) adj3 depressive symptomatology).ti,ab. (145);
42. IDS.ti,ab. (1174);
43. (Montgomery-Asberg adj3 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab. (894);
44. MADRS.ti,ab. (793);
45. (Depressive Adjective adj3 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab. (2);
46. DACL.ti,ab. (59);
47. (State-Trait anxiety adj3 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab. (1878);
48. STAL.ti,ab. (1219);
49. ((Depress\$ or anxiety) adj3 (Inventory or Questionnaire or scale or index or checklist or interview)).ti,ab. (23098);
50. QUESTIONNAIRES/(188001);
51. interview/(19930);
52. or/22–51 (4855611);
53. (comment or editorial or note).pt. (563657);
54. ((Screen\$ or diagnos\$ or predict\$ or detect\$ or aware\$) adj4 depress\$).ti,ab. (13746);
55. ((Screen\$ or diagnos\$ or predict\$ or detect\$ or aware\$) adj4 anxiety).ti,ab. (3666);
56. ((Screen\$ or diagnos\$ or predict\$ or detect\$ or aware\$) adj4 anxious).ti,ab. (203);
- 57.8 and 21 and 52 (448);
58. or/54–56 (16582);
59. 58 and 21 (114);
60. (57 or 59) not 53 (516).

References

- Andersen, H.S., 2004. Mental health in prison populations: a review—with special emphasis on a study of Danish prisoners on remand. *Acta Psychiatr. Scand.* 110, 5–59.
- BMA and NHS Employers, 2006. Revisions to the GMS contract, 2006/7. Delivering Investment in General Practice. British Medical Association, London.
- Brooke, D., Taylor, C., Gunn, J., Maden, A., 1996. Point prevalence of mental disorder in unconvicted male prisoners in England and Wales. *Br. Med. J.* 313, 1524–1527.
- Centre for Reviews and Dissemination, 2009. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care. University of York, York.
- Cooper, C., Berwick, S., 2001. Factors affecting psychological well-being of three groups of suicide-prone prisoners. *Current Psychology* 20, 169–182.
- Deville, W.L., Buntinx, F., Bouter, L.M., Montori, V.M., de Vet, H.C., van der Windt, D.A., Bezemer, P.D., 2002. Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med. Res. Methodol.* 2, 9.
- Duncan, A., Sacks, S., Cleland, C., Pearson, F., Coen, C., 2008. Performance of the CJDATS Co-Occurring Disorders Screening Instruments (CODSIs) among minority offenders. *Behav. Sci. Law* 26, 351–368.
- Gilbody, S., Richards, D., Brealey, S., Hewitt, C., 2007. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *J. Gen. Intern. Med.* 22, 1596–1602.
- Harrison, K.S., Rogers, R., 2007. Axis I screens and suicide risk in jails a comparative analysis. *Assessment* 14, 171–180.
- Hurley, W., Dunne, M.P., 1991. Psychological distress and psychiatric morbidity in women prisoners. *Aust. NZ J. Psychiatry* 25, 461–470.
- Kroenke, K., Spitzer, R.L., Williams, J.B., 2001. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* 16, 606–613.
- Kuo, E.S., Stoep, A.V., Stewart, D.G., 2005. Using the short mood and feelings questionnaire to detect depression in detained adolescents. *Assessment* 12, 374–383.
- Lorr, M., Sonn, T.M., Katz, M.M., 1967. Toward a definition of depression. *Arch. Gen. Psychiatry* 17, 183–186.
- Mann, R., Hewitt, C.E., Gilbody, S.M., 2008. Assessing the quality of diagnostic studies using psychometric instruments: applying QUADAS. *Soc. Psychiatry Psychiatr. Epidemiol.* 44, 300–307.
- McLearn, A.M., Ryba, N.L., 2003. Identifying severely mentally ill inmates: can small jails comply with detection standards? *Journal of Offender Rehabilitation* 37, 25–40.
- McMillan, D., Gilbody, S. and Beresford, E. (2007) Can we predict suicide and non-fatal self-harm with the Beck Hopelessness Scale? A meta-analysis. *Psychological Medicine* 37, 769–778.
- Mitchell, A.J., Coyne, J.C., 2007. Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. *Br. J. Gen. Pract.* 57, 144–151.
- Murray, C.J., Lopez, A.D., 1996. The Global Burden of Disease: a Comprehensive Assessment of Mortality and Disability from Disease, Injuries and Risk Factors in 1990. Harvard School of Public Health on behalf of the World Bank, Boston Mass.
- Nicholls, T., Lee, Z., Corrado, R., Ogloff, J., 2004. Women inmates' mental health needs: evidence of the validity of the Jail Screening Assessment Tool (JSAT). *International Journal of Forensic Mental Health* 3, 167–184.
- Perry, A.E., Gilbody, S., 2009. Detecting and predicting self-harm behaviour in prisoners: a prospective psychometric analysis of three instruments. *Soc. Psychiatry Psychiatr. Epidemiol.* 44, 853–861.
- Sacks, S., Melnick, G., Coen, C., 2007a. CJDATS co-occurring disorders screening instrument for mental disorders: a validation study. *Criminal Justice and Behavior* 34, 1198–1215.
- Sacks, S., Melnick, G., Coen, C., Banks, S., Friedmann, P., Grella, C., Knight, K., 2007b. CJDATS co-occurring disorders screening instrument for mental disorders: a pilot study. *The Prison Journal* 87, 86–109.
- Shaw, J., Tomenson, B., Creed, F., 2003. A screening questionnaire for the detection of serious mental illness in the criminal justice system. *Journal of Forensic Psychiatry & Psychology* 14, 138–150.
- Shaw, J., Baker, D., Hunt, I.M., Moloney, A., Appleby, L., 2004. Suicide by prisoners national clinical survey. *Br. J. Psychiatry* 184, 263–267.
- Spitzer, R.L., Kroenke, K., Williams, J.B.W., 1999. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Eds, *Am Med Assoc*, Vol. 282, pp. 1737–1744.
- Steadman, H., Scott, J., Osher, F., Agnese, T., Robbins, P., 2005. Validation of the brief jail mental health screen. *Psychiatr. Serv.* 56, 816–822.
- Steadman, H.J., Robbins, P.C., Islam, T., Osher, F.C., 2007. Revalidating the brief jail mental health screen to increase accuracy for women. *Psychiatr. Serv.* 58, 1598–1601.
- Teplin, L., Swartz, J., 1989. Screening for severe mental disorder in jails: the development of the referral decision scale. *Law Hum. Behav.* 13, 1–18.
- Vedavanam, S., Steel, N., Broadbent, J., Maisey, S., Howe, A., 2009. Recorded quality of care for depression in general practice: an observational study. *Br. J. Gen. Pract.* 59, e32–e37.
- Whiting, P., Rutjes, A.W.S., Reitsma, J.B., Glas, A.S., Bossuyt, P.M.M., Kleijnen, J., 2004. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann. Intern. Med.* 140, 189–202.
- Williams, J.W., Pignone, M., Ramirez, G., Stellato, C.P., 2002. Identifying depression in primary care: a literature synthesis of case-finding instruments. *Gen. Hosp. Psychiatry* 24, 225–237.
- Youden, W.J., 1950. Index for rating diagnostic tests. *Cancer* 3, 32–35.