

Accuracy of specific symptoms in the diagnosis of major depressive disorder in psychiatric out-patients: data from the MIDAS project

A. J. Mitchell¹*, J. B. McGlinchey², D. Young², I. Chelminski² and M. Zimmerman²

¹ Liaison Psychiatry, Leicester General Hospital and Department of Cancer and Molecular Medicine, Leicester Royal Infirmary, Leicester, UK

² Department of Psychiatry and Human Behavior, Brown University School of Medicine, Rhode Island Hospital, Providence, RI, USA

Background. There is uncertainty about the diagnostic significance of specific symptoms of major depressive disorder (MDD). There is also interest in using one or two specific symptoms in the development of brief scales. Our aim was to elucidate the best possible specific symptoms that would assist in ruling in or ruling out a major depressive episode in a psychiatric out-patient setting.

Method. A total of 1523 psychiatric out-patients were evaluated in the Methods to Improve Diagnostic Assessment and Services (MIDAS) project. The accuracy and added value of specific symptoms from a comprehensive item bank were compared against the Structured Clinical Interview for DSM-IV (SCID).

Results. The prevalence of depression in our sample was 54.4%. In this high prevalence setting the optimum specific symptoms for ruling in MDD were psychomotor retardation, diminished interest/pleasure and indecisiveness. The optimum specific symptoms for ruling out MDD were the absence of depressed mood, the absence of diminished drive and the absence of loss of energy. However, some discriminatory items were relatively uncommon. Correcting for frequency, the most clinically valuable rule-in items were depressed mood, diminished interest/pleasure and diminished drive. The most clinically valuable rule-out items were depressed mood, diminished interest/pleasure and poor concentration.

Conclusions. The study supports the use of the questions endorsed by the two-item Patient Health Questionnaire (PHQ-2) with the additional consideration of the item diminished drive as a rule-in test and poor concentration as a rule-out test. The accuracy of these questions may be different in primary care studies where prevalence differs and when they are combined into multi-question tests or algorithmic models.

Received 25 September 2007; Revised 3 September 2008; Accepted 16 September 2008; First published online 12 November 2008

Key words: Clinical utility, diagnostic accuracy, major depression, sensitivity, specificity.

Introduction

Major depressive disorder (MDD) is typically a relapsing remitting illness that rarely occurs as a solitary episode. Even depressions treated in primary care tend to be recurrent, chronic and co-morbid (Judd, 1997; Lin *et al.* 1998; Brodaty *et al.* 2001; Gilmer *et al.* 2005; Vuorilehto *et al.* 2005). Numerous publications have drawn attention to the low detection of MDD in primary care with a typical case recognition rate (sensitivity of unassisted clinical detection alone) of between 36% and 56% (Thompson *et al.* 2000; Christensen *et al.* 2003; Croudace *et al.* 2003; MaGPIe Research Group, 2004). An equivalent body of work

has highlighted low detection rates in medical settings (Wilhelm *et al.* 2004). Little, however, has been published regarding the detection rates in psychiatric settings or indeed about the assessment and screening practices of psychiatrists. Most information concerning the diagnostic value of specific symptoms of depression comes from validation studies of various mood questionnaires, including those replying upon only one or two questions (Williams *et al.* 2002*a,b*; Takeuchi *et al.* 2006; Li *et al.* 2007; Mitchell & Coyne, 2007). Yet this approach may be problematic because almost every scale and diagnostic schedule was created by consensus without primary data regarding the value of specific symptoms suggested. Furthermore, although there is considerable evidence concerning the accuracy of scores generated from combining multiple questions for depression, there is very little evidence concerning the diagnostic value of individual symptoms, even those that are included in DSM-IV

* Address for correspondence: A. J. Mitchell, MBBS, Consultant in Liaison Psychiatry, Leicester General Hospital, Leicester LE5 4PW, UK.

(Email: alex.mitchell@leicspart.nhs.uk)

and ICD-10. In ICD-10 two typical symptoms are required from the following three items: depressed mood, loss of interest, and decreased energy. A minimum of four symptoms are required to qualify with mild depression, and five symptoms (later revised to six) needed for moderate depression. To qualify as a severe depressive episode all three typical symptoms must be present plus at least four other symptoms. In DSM-IV either depressed mood or loss of interest is required for a diagnosis of MDD, with a total of five of a list of nine symptoms altogether. In theory, assigning special significance to core features reduces false-positive diagnoses in those patients who manifest five of the nine criteria but without low mood or lost interest. Nevertheless, Zimmerman *et al.* (2006a) found that only 27 (1.5%) of 1800 psychiatric outpatients reported five or more criteria *in the absence* of low mood or loss of interest or pleasure. Of these 27 patients, 25 reported depressed mood at a sub-threshold level. It is therefore unclear whether low mood or loss of interest have special significance in relation to a diagnosis of MDD. Indeed, it is also unclear to what degree all proposed specific symptoms of MDD have diagnostic weight when considered on their own.

The accuracy of specific symptoms may have additional importance in relation to screening for depression. There has been interest in developing very brief screening tools with less than five questions in the hope of improving acceptability in primary care (Takeuchi *et al.* 2006; Mitchell & Coyne, 2007; Muhwezi *et al.* 2007). Although the nine DSM-IV criteria of MDD have been incorporated into a short instrument, the Patient Health Questionnaire (PHQ-9), in one survey 62.5% of general practitioners considered this questionnaire too long and 37.5% considered it too time-consuming (Bermejo *et al.* 2005). This has led to the development of ultra-short questionnaires consisting of two or three questions, or even just a single detection question. Perhaps the most well-known example is the two-item Patient Health Questionnaire (PHQ-2) (Spitzer *et al.* 1999). This asks: Over the past 2 weeks, have you been bothered by either (a) feeling down, depressed or hopeless or (b) having little interest or pleasure in doing things? From early validation studies on the PHQ it is likely that the items low mood (strictly a three-part conjoint question) and loss of interest (strictly a two-part conjoint question) were selected because they were the essential features in DSM-IV. In 2004 the National Institute for Health and Clinical Excellence (NICE) released guidelines for the management of unipolar depression in primary and secondary care (NICE, 2004). In 2007 NICE released guideline 45 for antenatal and postnatal mental health (NICE, 2007). Both included

the recommendation that simple screening using two or three questions would suffice, and offering an adapted version the PHQ-2 that extended over a duration of 4 weeks rather than 2. Although the PHQ-2 has been used in a range of studies, only two studies have reported the accuracy of the questions applied individually (head-to-head) (Whooley *et al.* 1997; Lowe *et al.* 2003). In both of these studies, the second PHQ question (loss of interest) alone had superior sensitivity (Se) and negative predictive value (NPV) to the first question (low mood). It remains untested whether these items are optimal for diagnosing MDD or whether other specific symptoms would be preferable. To our knowledge, no group has reported on the diagnostic validity of the PHQ-2 methods in specialist settings although one publication reported on the PHQ-9 in 171 'psychosomatic out-patients' (Graf *et al.* 2004). Similarly, no group has attempted to examine the diagnostic value of specific symptoms in psychiatric settings.

The Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project is a large clinical epidemiological study in which semi-structured interviews were administered to a large sample of patients presenting for psychiatric outpatient treatment. We have previously examined the diagnostic properties of the DSM-IV criteria, in addition to the psychometric performance of symptoms that are not part of the diagnostic criteria (McGlinchey *et al.* 2006). Using simple logistic regression we found that the ranked order of symptoms by diagnostic weight for DSM-IV membership was depressed mood > loss of interest (anhedonia) > sleep disturbance > concentration/indecision > worthlessness/excessive guilt > loss of energy (Zimmerman *et al.* 2006b). The aim of the current study was to re-examine the diagnostic validity of a full item bank of DSM-IV and non-DSM-IV symptoms in order to determine which single items would be the most useful as a single-item diagnostic rule-in or rule-out test for MDD, as applied to a high prevalence setting.

Method

To date in the MIDAS project 1800 psychiatric outpatients have been evaluated with a semi-structured diagnostic interview in the Rhode Island Hospital Department of Psychiatry outpatient practice. The methods of the study have been described in detail elsewhere (Zimmerman *et al.* 2006b). The Rhode Island Hospital institutional review committee approved the research protocol, and all patients provided informed, written consent. Patients were interviewed by a diagnostic rater who administered the Structured Clinical

Interview for DSM-IV (SCID; First *et al.* 1995). To study the psychometric performance of the DSM-IV symptom criteria for major depression, it was necessary to modify the SCID and eliminate the skip-out that curtails the depression module for patients who did not report either depressed mood or loss of interest or pleasure. Thus, we enquired about all of the symptoms of depression for all patients. For compound criteria that encompass more than one symptom (e.g. indecisiveness or impaired concentration; increased sleep or insomnia), we made separate ratings of each component of the diagnostic criterion. Thus, the nine DSM-IV symptom criteria were broken down into 17 separate items. Our total item bank consisted of 25 items (22 single and three dual or combination items). The combination items we allowed were 'diminished interest or pleasure', 'anxiety' and 'sleep disturbance', as many consider these to be integral features of depression without separation into more specific components.

Patients were interviewed by a diagnostic rater who administered the SCID, supplemented by items from the Schedule for Affective Disorders and Schizophrenia (SADS), to rate the severity of depressive and non-depressive symptoms (Endicott & Spitzer, 1978). The diagnostic raters in the MIDAS project were highly trained and monitored throughout the study to minimize rater drift. In addition to rating the DSM-IV MDD criteria, the interviewers determined the presence of the following symptoms, which are not part of the diagnostic criteria: hopelessness, helplessness, unreactive mood, diminished drive, psychic anxiety, and somatic anxiety. The reasons for considering these symptoms as possible diagnostic indicators for depression have been detailed previously (McGlinchey *et al.* 2006).

Out of the pool of 1800 patients, we excluded from the present analysis 106 patients with current bipolar disorder because there are some symptom differences between patients with bipolar and non-bipolar forms of depression. We also excluded 171 patients who had MDD that was in partial remission because inclusion of these patients with the depression group would have lowered the Se of the symptom criteria, whereas inclusion of the patients in the non-depressed group would have reduced specificity (Sp). Thus, the present report is based on the 1523 remaining psychiatric out-patients who were administered a semi-structured diagnostic interview. The demographic characteristics of the sample are included in Table 1. In brief, the majority of the subjects were white, female, married, and single. The most frequent DSM-IV diagnoses were MDD (46.0%, $n=829$), social phobia (28.8%, $n=519$), panic disorder (18.4%, $n=331$) and generalized anxiety disorder (17.8%, $n=320$).

Table 1. Demographic characteristics of sample ($n=1523$)

Characteristic	<i>n</i>	%
Gender		
Male	601	39.5
Female	922	60.5
Ethnicity		
Caucasian	1326	87.1
African-American	64	4.2
Portuguese	55	3.6
Hispanic	42	2.8
Asian	12	0.8
Other	24	1.6
Marital status		
Married	629	41.3
Never married	466	30.6
Divorced	211	13.9
Separated	95	6.2
Living together as if married	92	6.0
Widowed	30	2.0
Education		
Less than high school	178	11.7
Graduated high school	964	63.3
Graduated college or greater	381	25.0
Age (years), mean (s.d.)	38.0 (12.8)	

s.d., Standard deviation.

Diagnostic validity testing

In examining diagnostic accuracy of specific symptoms, a number of methodological issues were considered. The performance of a test will vary with the baseline prevalence of the condition (Whiting *et al.* 2004). Rule-in accuracy is computed as the product of the positive predictive value (PPV), where the denominator is all who test positive, and Sp, where the denominator is all without the disease (Sackett & Haynes, 2002). For example, if the Sp is 100%, then there can be no false positives and hence all positive scores will imply a true case. Rule-out accuracy is computed as the product of the NPV, where the denominator is all who test negative, and Se, where the denominator is all with the disease (Sackett & Haynes, 2002). When using a clinical feature as a diagnostic test, a symptom may have a high PPV or NPV, but in the real world its clinical relevance will be poor if it occurs rarely. Consider the example of a hypothetical biological challenge test that, if positive, has a 90% PPV but is only positive in half of depressed individuals (Se 50%). Clinically relevant rule-in accuracy would be the product of the PPV and Se and clinically relevant rule-out accuracy (rule-out accuracy corrected for occurrence) would be the product of the NPV and Sp. These have been defined as the positive utility index (UI+) and the negative utility index

Table 2. Psychometric performance of symptoms of depression in 1523 psychiatric out-patients evaluated with a semi-structured interview: measures of accuracy of all single-item symptoms of depression from the SCID, against DSM-IV

Item	Occurrence in depressed (<i>n</i> = 829)	Se	Non-occurrence in non-depressed (<i>n</i> = 694)	Sp	PPV	NPV	PSI	Youden's <i>J</i>
Depressed mood	770	0.93	572	0.82	0.86	0.91	0.77	0.75
Diminished drive	731	0.88	485	0.70	0.78	0.83	0.61	0.58
Loss of energy	723	0.87	475	0.68	0.77	0.82	0.59	0.56
Sleep disturbance (combined)	689	0.83	474	0.68	0.76	0.77	0.53	0.51
Diminished concentration	678	0.82	525	0.76	0.80	0.78	0.58	0.57
Diminished interest/pleasure	668	0.81	609	0.88	0.89	0.79	0.68	0.68
Sleep disturbance (insomnia)	582	0.70	506	0.73	0.76	0.67	0.43	0.43
Anxiety (combined)	572	0.69	401	0.58	0.66	0.61	0.27	0.27
Worthlessness	508	0.61	613	0.88	0.86	0.66	0.52	0.50
Helplessness	495	0.60	587	0.85	0.82	0.64	0.46	0.44
Anxiety (psychic anxiety)	487	0.59	468	0.67	0.68	0.58	0.26	0.26
Thoughts of death	463	0.56	608	0.88	0.84	0.62	0.47	0.43
Hopelessness	460	0.55	589	0.85	0.81	0.61	0.43	0.40
Excessive guilt	448	0.54	604	0.87	0.83	0.61	0.45	0.41
Indecisiveness	424	0.51	638	0.92	0.88	0.61	0.50	0.43
Anxiety (somatic anxiety)	380	0.46	520	0.75	0.69	0.54	0.22	0.21
Decreased appetite	370	0.45	620	0.89	0.83	0.57	0.41	0.34
Anger	361	0.44	513	0.74	0.67	0.52	0.19	0.17
Psychomotor agitation	286	0.34	631	0.91	0.82	0.54	0.36	0.25
Psychomotor retardation	232	0.28	668	0.96	0.90	0.53	0.43	0.24
Decreased weight	189	0.23	650	0.94	0.81	0.50	0.32	0.16
Lack of reactive mood	179	0.22	650	0.94	0.80	0.50	0.30	0.15
Sleep disturbance (hypersomnia)	160	0.19	652	0.94	0.79	0.49	0.29	0.13
Increased appetite	157	0.19	642	0.93	0.75	0.49	0.24	0.11
Increased weight	131	0.16	650	0.94	0.75	0.48	0.23	0.09

Se, Sensitivity; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value; PSI, Predictive Summary Index.

(UI-) respectively (Mitchell, 2008). Several methods are available to calculate overall diagnostic accuracy. The pre-test post-test gain, or 'added value', may be calculated by the difference between the prevalence and either PPV or NPV. Summary methods of accuracy include Youden's *J* and the Predictive Summary Index (PSI; Youden, 1950). Youden's *J* is a composite of overall accuracy using $Se + Sp - 1$. The PSI is a composite of overall accuracy using all positive and negative screens calculated as: $PPV + NPV - 1$. Where multiple tests generate different Se and Sp values, the results can be combined in a summary receiver operating characteristic (sROC) curve (Macaskill, 2004).

In the analysis of specific symptoms there is no cut-off, as all items score categorically. A further issue is that there is no consensus about what level of accuracy is acceptable clinically. The acceptable ratio of Se/Sp may depend on how important it is not to overlook cases (false negatives) or incorrectly diagnose cases (false positives). In some situations it may be better to minimize false negatives at the expense of false

positives. As a guide we used the following grades of diagnostic accuracy (applied to PPV or NPV) adapted from Landis & Koch (1977): excellent = 0.90, good = 0.80, satisfactory = 0.80; otherwise poor. The equivalent grades for clinical utility were 0.81, 0.64 and 0.49 (applied to the utility index).

Results

The sample characteristics are shown in Table 1.

Symptom frequencies

Seven items occurred in at least half of all cases (depressed and non-depressed). The all-case proportion was, in diminishing order: loss of energy (62%), diminished drive (62%), sleep disturbance (60%) depressed mood (59%), anxiety (57%), diminished concentration (56%), and insomnia (51%). All of these items also occurred in more than half of patients with MDD. The only items that were common (> 50%)

Table 3. Five most and five least successful rule-in items

	PPV	Added value (%) (PPV-prevalence)	Occurrence in depressed (%)	Grade of accuracy ^a	Grade of clinical utility ^b
Five most successful rule-in items					
1. Psychomotor retardation	0.90	35.5	27.9	Excellent	Poor
2. Diminished interest/pleasure	0.89	34.3	81.0	Good	Good
3. Indecisiveness	0.88	33.9	51.1	Good	Poor
4. Depressed mood	0.86	31.9	92.8	Good	Good
5. Worthlessness	0.86	31.8	61.3	Good	Satisfactory
Five least successful rule-in items					
5. Increased weight	0.74	20.4	15.8	Poor	Poor
4. Somatic anxiety	0.69	14.2	36.4	Poor	Poor
3. Psychic anxiety	0.68	13.9	46.8	Poor	Poor
2. Anger	0.67	12.2	35.6	Poor	Poor
1. Anxiety	0.66	11.7	56.8	Poor	Poor

^a Refers to positive predictive value (PPV) in this case.

^b Refers to positive utility index.

in MDD but uncommon in non-MDD were diminished interest/pleasure, helplessness, appetite/weight disturbance, indecisiveness and psychomotor retardation.

The 10 most common symptoms in patients with depression were: depressed mood (93%), diminished drive (88%), loss of energy (87%), sleep disturbance (83%), diminished concentration (82%), diminished interest/pleasure (81%), insomnia (70%), anxiety (69%), worthlessness (61%) and helplessness (60%).

Rule-in accuracy

Overall rule-in accuracy

The Se, Sp, PPV and NPV of all items considered are shown in Table 2. Using the PPV, the five most accurate symptoms and the five least accurate symptoms when confirming a diagnosis of MDD are shown in Table 3. From these results, in a psychiatric out-patient setting, if a patient had psychomotor retardation, there is a 90% chance based on this item alone that MDD would be present. However, the presence of anxiety alone increases the chance of a correct diagnosis by only 11.7% above the baseline (chance) rate (calculated by subtracting prevalence from PPV). In fact, psychomotor retardation only occurred in 16.9% of the sample and 28% of depressed patients, reducing its clinical applicability.

Rule-in accuracy corrected for occurrence (UI)

Considering the UI+, that is the occurrence in the depressed patients (Se) multiplied by the proportion of all positive tests that are accurate (PPV), the top five most accurate and frequent items were: depressed

mood (UI+0.80), diminished interest/pleasure (UI+0.71), diminished drive (UI+0.69), loss of energy (UI+0.67) and diminished concentration (UI+0.65).

Clinically, out of 200 individuals seen as psychiatric out-patients where the prevalence of depression is approximately 50%, 111 would be expected to have depressed mood and, of these, 93 would be true cases and 18 false positives. Ninety-one individuals would be expected to deny depressed mood and, of these, 82 would be true negatives and seven with syndromal depressed overlooked.

Rule-out accuracy

Overall rule-out accuracy

Using the NPV, the five most accurate and the five least accurate symptoms when attempting to confirm the absence of MDD are shown in Table 4. Thus, in a clinical setting, if an individual did not report depressed mood, there was a 91% chance that the individual was not suffering depression, an added value of 36.2% over chance detection alone. However, if an individual did not report weight loss, there was only a 50% chance based on this item alone that MDD would be absent. This is in fact 4% less than the unassisted detection rate of 54.4%, suggesting that this item is less than helpful in ruling-out depression. A summary of the added value of all specific symptoms is illustrated in Fig. 1.

Rule-out accuracy correct for occurrence (UI)

Considering the UI-, or the occurrence in the non-depressed patients (Sp) in combination with the proportion of all negatives tests that are accurate

Table 4. Five most and five least successful rule-out items

	NPV	Added value (%) (NPV-prevalence)	Non-occurrence in non-depressed (%)	Grade of accuracy ^a	Grade of clinical utility ^b
Five most successful rule-out items					
1. Depressed mood	0.91	36.2	82.4	Excellent	Good
2. Diminished drive	0.83	28.7	70.0	Good	Satisfactory
3. Loss of energy	0.82	27.3	68.4	Good	Satisfactory
4. Diminished interest/pleasure	0.79	24.7	87.7	Good	Good
5. Diminished concentration	0.78	23.2	75.6	Good	Satisfactory
Five least successful rule-out items					
5. Decreased weight	0.50	-4.0	93.6	Poor	Poor
4. Lack of reactive mood	0.50	-4.4	93.6	Poor	Poor
3. Hypersomnia	0.49	-5.1	94.0	Poor	Poor
2. Increased appetite	0.49	-5.6	92.5	Poor	Poor
1. Increased weight	0.48	-6.2	93.6	Poor	Poor

^a Refers to negative predictive value (NPV) in this case.

^b Refers to negative utility index.

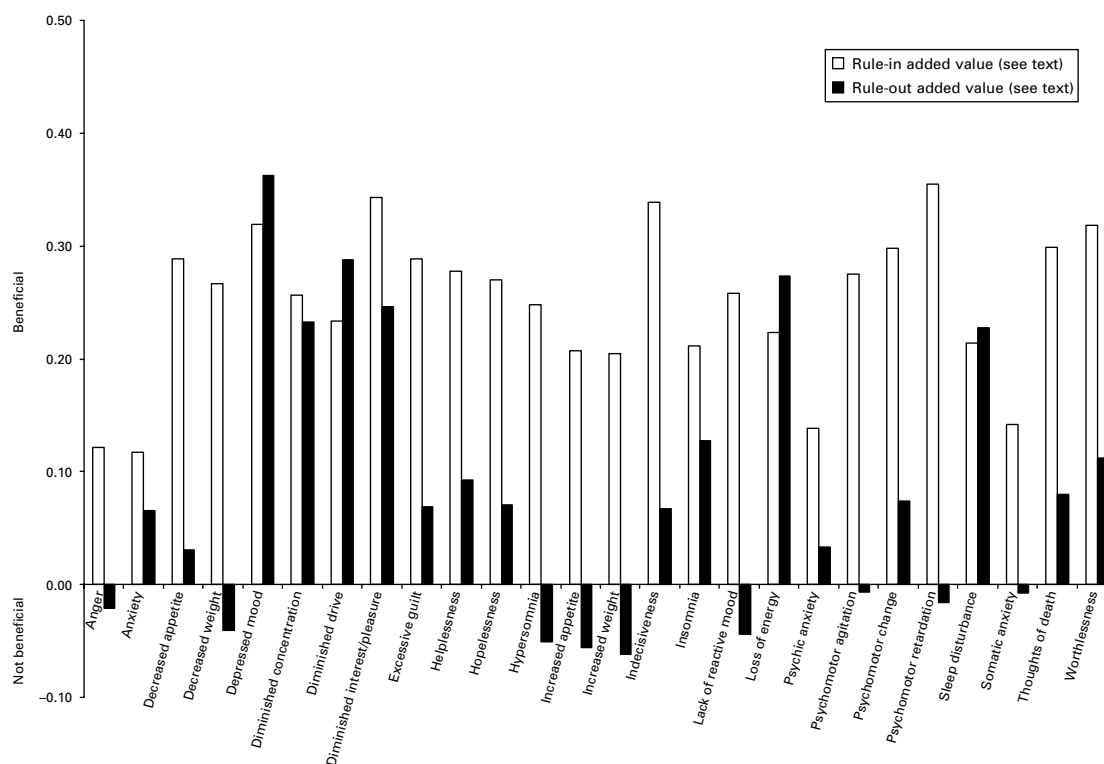


Fig. 1. Added value of specific symptoms in diagnosing major depression.

(NPV), the top five most accurate and frequent items were: depressed mood (UI- 0.75), diminished interest/pleasure (UI- 0.69), diminished concentration (UI- 0.59), diminished drive (UI- 0.58) and worthlessness (UI- 0.57).

Using the item diminished interest/pleasure in a hypothetical group of 200 out-patients in which the prevalence of depression is about 50%, 92 people would be expected to show loss of interest/pleasure and 108 would not. Of these 92, 80 (88%) would be

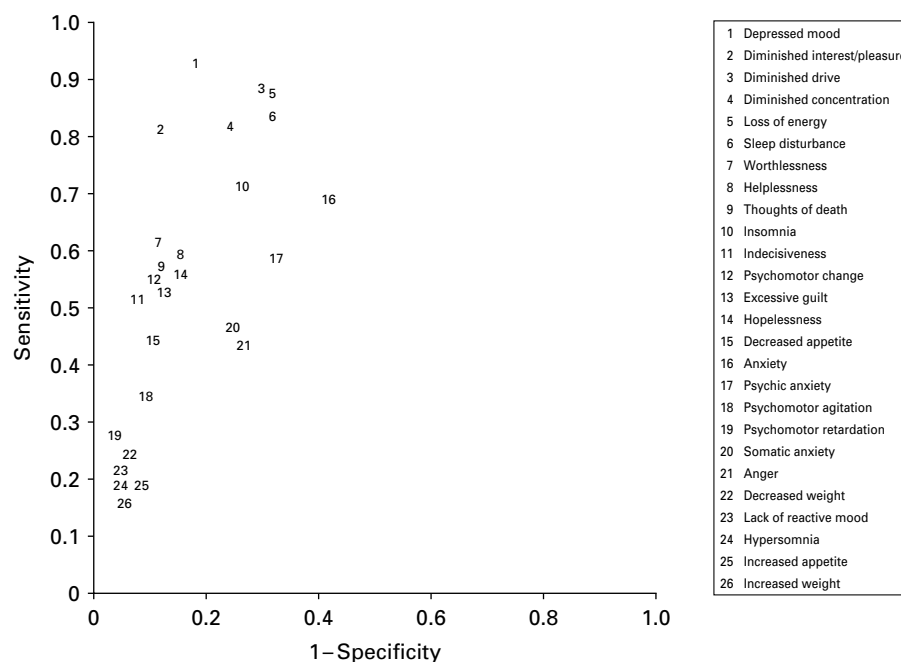


Fig. 2. Summary receiver operating characteristic (sROC) curve plot of accuracy of individual mood symptoms in the diagnosis of major depressive disorder (MDD).

true positives and of the 108 without this symptom, 88 (79%) would be true negatives.

Combined accuracy

Measures of combined accuracy attempt to assess overall diagnostic accuracy considering rule-in and a rule-out test at the same time. Examining the Youden index and the PSI, the most discriminating items were: depressed mood (Youden=0.75), diminished interest/pleasure (Youden=0.68) and diminished drive (Youden=0.58). This is illustrated in a sROC plot of Se versus 1 – Sp for each individual item (Fig. 2).

Discussion

In this report from the MIDAS project, we extended our previous work on discriminatory items by looking in more detail at rule-in and rule-out accuracy of specific symptoms of depression. As anticipated, the most common symptom of depression in this sample was depressed mood but, of note, the next four most common (diminished drive, loss of energy, sleep disturbance, diminished concentration) are traditionally viewed as somatic or cognitive rather than affective in nature. These findings correspond to previous studies that have examined the frequency of symptoms of depression in psychiatric, primary care and community samples (Breslau & Davis, 1985; Buchwald & Rudick-Davis, 1993). However, frequency of occurrence

is not the same as discriminatory ability and this in turn is best considered in two directions. In this sample the three most discriminatory items for ruling in MDD were psychomotor retardation, diminished interest/pleasure and indecisiveness. The three most discriminatory specific symptoms for ruling out MDD were the absence of depressed mood, diminished drive and loss of energy. However, some rule-in items were uncommon in depressed patients and some rule-out items were uncommon in non-depressed patients so that their usefulness in the clinical setting would be limited. Correcting for this, most clinically useful rule-in items became depressed mood, diminished interest/pleasure and diminished drive. The most clinically useful rule-out items became the absence of depressed mood, the absence of diminished interest/pleasure and the absence of poor concentration. The combined measures of accuracy ranked discriminatory power as depressed mood > diminished interest/pleasure > diminished drive. These items increased the chances of an accurate diagnosis from about 50% to more than 80% even when used alone.

The results suggest that the low mood and loss of interest/pleasure items of the PHQ-2 and NICE guidance are close to the optimal single-item questions out of all possible symptoms of depression. The fact that the best single item was depressed mood is notable, given that Zimmerman *et al.* (2006c) recently found that only 7% of those qualifying as MDD do not report low mood. Patients without reported depressed mood

may be qualitatively different to those with low mood and hence may be more problematic for general practitioners to detect. Diminished drive had significant value in diagnosing syndromal depression. Diminished drive refers to reduced motivation or a decreased capacity to initiate behaviour towards a certain goal. Further work should be performed to elucidate the role of motivation as a diagnostic feature of MDD. Loss of energy (fatigue), a typical (core) feature of ICD-10, had only modest discriminatory power, in part because it occurred in 32% of non-depressed cases. However, somatic symptoms were common and important. Most previous studies agree that somatic symptoms of depression are common and discriminating in both primary and secondary care settings (Akechi *et al.* 2003; Nakao & Yano, 2003; Barkow *et al.* 2004; Reuter *et al.* 2004; de Coster *et al.* 2005). However, in a study of Nigerian army personnel using principal component analysis, Okulate *et al.* (2004) found that somatic items accounted for only a little of the total variance for depression in this setting. Patients who present initially with somatic complaints may pose diagnostic difficulties if health-care professionals only think about depression when emotional complaints are mentioned (Bridges & Goldberg, 1985).

The strengths of this study are the large sample size and the use of highly trained, reliable interviewers who used semi-structured diagnostic interviews. There are also several limitations. First, the alternative (non-DSM-IV) symptoms may have been disadvantaged, in that a diagnosis of MDD was dependent on the contribution of each of the DSM-IV MDD symptom criteria but not influenced at all by the alternative symptoms. Second, data on diagnostic accuracy were calculated *post hoc*. Dissecting individual items from a questionnaire may increase the discriminatory ability compared with a new independent analysis because an interviewer is not required to use all items to make a diagnosis. Third, and perhaps most importantly, we relied on a psychiatric out-patient sample where the prevalence of depression was 54.4%. This is considerably higher than that seen in studies of MDD in primary care. This may mean that the diagnostic weighting of individual symptoms is somewhat different in a primary care sample. This clearly requires further study.

Miller (2002) found that, when unassisted, clinicians evaluated an average of only 32% of DSM-IV criteria for MDD. Even psychiatrists, who usually remember to ask about low mood, enquire about loss of interest/pleasure in only 8% of evaluations for depression (Miller, 2002). General practitioners only consider low mood or loss of interest to be useful in detecting depression in 54% and 36% of cases respectively

(Krupinski & Tiller, 2001). Given these problems, considerable effort has been expended on perfecting short case-finding instruments that might perform almost as well as longer validated severity scales or even semi-structured interviews. In a large occupational health sample of 1621 workers, Takeuchi *et al.* (2006) reported that the single item 'feeling blue' or a combination with 'miserable' had similar diagnostic accuracy to the full 15-item Profile of Mood States (POMS) questionnaire. Mitchell & Coyne (2007) recently reported pooled analysis from eight studies of single-question tests for the diagnosis of depression in primary care. Although the overall Se was low at 31.9%, Sp was high at 96% (PPV was 55.6% and NPV was 92.3%).

In conclusion, we found that the most clinically valuable specific symptoms when diagnosing depression were depressed mood and diminished interest/pleasure together with diminished drive (rule-in) and poor concentration (rule-out). This study supports the use of the questions endorsed by the PHQ-2 in a psychiatric out-patient clinic and suggests that specific items can give reasonable diagnostic performance in high prevalence settings, providing that 14% false-positive and 10% false-negative error rates are considered acceptable.

Declaration of Interest

None.

References

- Akechi T, Nakano T, Akizuki N, Okamura M, Sakuma K, Nakanishi T, Yoshikawa E, Uchitomi Y (2003). Somatic symptoms for diagnosing major depression in cancer patients. *Psychosomatics* **44**, 244–248.
- Barkow K, Heun R, Ustun TB, Berger M, Bermejo I, Gaebel W, Harter W, Schneider F, Stieglitz RD, Maier W (2004). Identification of somatic and anxiety symptoms which contribute to the detection of depression in primary health care. *European Psychiatry* **19**, 250–257.
- Bermejo I, Niebling W, Mathias B, Harter M (2005). Patients' and physicians' evaluation of the PHQ-D for depression screening. *Primary Care and Community Psychiatry* **10**, 125–131.
- Breslau N, Davis G (1985). Refining DSM-III criteria in major depression: an assessment of the descriptive validity of criterion symptoms. *Journal of Affective Disorders* **9**, 199–206.
- Bridges KW, Goldberg DP (1985). Somatic presentation of DSM-III psychiatric disorders in primary care. *Journal of Psychosomatic Research* **29**, 563–569.
- Brodaty H, Luscombe G, Peisah C, Anstey K, Andrews G (2001). A 25-year longitudinal, comparison study of the outcome of depression. *Psychological Medicine* **31**, 1347–1359.

- Buchwald A, Rudick-Davis D** (1993). The symptoms of major depression. *Journal of Abnormal Psychology* **102**, 197–205.
- Christensen KS, Toft T, Frostholm L, Ørnbøl E, Fink P, Olesen F** (2003). The FIP Study: a randomised, controlled trials of screening and recognition of psychiatric disorders. *British Journal of General Practice* **53**, 758–763.
- Croudace T, Evans J, Harrison G, Sharp DJ, Wilkinson E, McCann G, Spence M, Crilly C, Brindle L** (2003). Impact of the ICD-10 Primary Health Care (PHC) diagnostic and management guidelines for mental disorders on detection and outcome in primary care. Cluster randomised controlled trial. *British Journal of Psychiatry* **182**, 20–30.
- de Coster L, Leentjens AFG, Lodder J, Verhey FRJ** (2005). The sensitivity of somatic symptoms in post-stroke depression: a discriminant analytic approach. *International Journal of Geriatric Psychiatry* **20**, 358–362.
- Endicott J, Spitzer RL** (1978). A diagnostic interview: the schedule for affective disorders and schizophrenia. *Archives of General Psychiatry* **35**, 837–844.
- First MB, Spitzer RL, Gibbon M, Williams JBW** (1995). *Structured Clinical Interview for DSM-IV Axis I Disorders – Patient Edition (SCID-I/P, Version 2.0)*. Biometrics Research Department, New York State Psychiatric Institute: New York.
- Gilmer WS, Trivedi MH, Rush AJ, Wisniewski SR, Luther J, Howland RH, Yohanna D, Khan A, Alpert J** (2005). Factors associated with chronic depressive episodes: a preliminary report from the STAR-D project. *Acta Psychiatrica Scandinavica* **112**, 425–433.
- Grafe K, Zipfel S, Herzog W, Lowe B** (2004). Screening for psychiatric disorders with the Patient Health Questionnaire (PHQ). Results from the German validation study. *Diagnostica* **50**, 171–181.
- Judd LL** (1997). The clinical course of unipolar major depressive disorders. *Archives of General Psychiatry* **54**, 989–991.
- Krupinski J, Tiller J** (2001). The identification and treatment of depression by general practitioners. *Australian and New Zealand Journal of Psychiatry* **35**, 827–832.
- Landis JR, Koch GG** (1977). The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174.
- Li CY, Friedman B, Conwell Y, Fiscella K** (2007). Validity of the Patient Health Questionnaire 2 (PHQ-2) in identifying major depression in older people. *Journal of the American Geriatrics Society* **55**, 596–602.
- Lin EH, Katon WJ, Von Korff M, Russo JE, Simon GE, Bush TM, Rutter CM, Walker EA, Ludman E** (1998). Relapse of depression in primary care: rate and clinical predictors. *Archives of Family Medicine* **7**, 443–449.
- Lowe B, Grafe K, Kroenke K, Zipfel S, Quenter A, Wild B, Fiehn C, Herzog W** (2003). Predictors of psychiatric comorbidity in medical outpatients. *Psychosomatic Medicine* **65**, 764–770.
- Macaskill P** (2004). Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. *Journal of Clinical Epidemiology* **57**, 925–932.
- MaGPIe Research Group** (2004). General practitioner recognition of mental illness in the absence of a ‘gold standard’. The Mental Health and General Practice Investigation (MaGPIe) Research Group. *Australian and New Zealand Journal of Psychiatry* **38**, 789–794.
- McGlinchey JB, Zimmerman M, Young D, Chelminski I** (2006). Diagnosing major depressive disorder VIII: are some symptoms better than others? *Journal of Nervous and Mental Disease* **194**, 785–790.
- Miller PR** (2002). Inpatient diagnostic assessments: 3. Causes and effects of diagnostic imprecision. *Psychiatry Research* **111**, 191–197.
- Mitchell AJ, Coyne JC** (2007). Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. *British Journal of General Practice* **57**, 144–151.
- Mitchell AJ** (2008). The clinical significance of subjective memory complaints in the diagnosis of mild cognitive impairment and dementia: a meta-analysis. *International Journal of Geriatric Psychiatry*. Published online: 23 May 2008. doi:10.1002/gps.2053.
- Muhwezi WW, Agren H, Musisi S** (2007). Detection of major depression in Ugandan primary health care settings using simple questions from a subjective well-being (SWB) subscale. *Social Psychiatry and Psychiatric Epidemiology* **42**, 61–69.
- Nakao M, Yano E** (2003). Reporting of somatic symptoms as a screening marker for detecting major depression in a population of Japanese white-collar workers. *Journal of Clinical Epidemiology* **56**, 1021–1026.
- NICE** (2004). *Depression: Management of Depression in Primary and Secondary Care*. Clinical Guideline 23. National Institute for Health and Clinical Excellence: London.
- NICE** (2007). *Antenatal and Postnatal Mental Health: Clinical Management and Service Guidance*. Clinical Guideline 45. National Institute for Health and Clinical Excellence: London.
- Okulate GT, Olayinka MO, Jones OBE** (2004). Somatic symptoms in depression: evaluation of their diagnostic weight in an African setting. *British Journal of Psychiatry* **184**, 422–427.
- Reuter K, Raugust S, Bengel J, Harter M** (2004). Depressive symptom patterns and their consequences for diagnosis of affective disorders in cancer patients. *Supportive Care in Cancer* **12**, 864–870.
- Sackett DL, Haynes RB** (2002). The architecture of diagnostic research. *British Medical Journal* **324**, 539–541.
- Spitzer RL, Kroenke K, Williams JB** (1999). Validation and utility of a self-report version of the PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *Journal of the American Medical Association* **282**, 1737–1744.
- Takeuchi T, Nakao M, Yano E** (2006). Screening for major depression in the workplace: testing diagnostic accuracy of a two-item questionnaire used during mandatory testing. *Primary Care and Community Psychiatry* **11**, 13–19.
- Thompson C, Kinmonth J, Stevens L, Peveler RC, Stevens A, Ostler KJ, Pickering RM, Baker N, Henson A, Preece J, Cooper D, Campbell MJ** (2000). Effects of a clinical-practice guideline and practice-based education on detection and outcome of depression in primary care: Hampshire Depression Project randomised controlled trial. *Lancet* **355**, 50–57.

- Vuorilehto M, Melartin T, Isometsa E** (2005). Depressive disorders in primary care: recurrent, chronic, and co-morbid. *Psychological Medicine* **35**, 673–682.
- Whiting P, Rutjes AWS, Dinnes J, Reitsma JB, Bossuyt PMM, Kleijnen J** (2004). Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technology Assessment* **8**, 1–234.
- Whooley MA, Avins AL, Miranda J, Browner WS** (1997). Case-finding instruments for depression. Two questions are as good as many. *Journal of General and Internal Medicine* **12**, 439–445.
- Wilhelm K, Kotze B, Waterhouse M, Hadzi-Pavlovic D, Parker G** (2004). Screening for depression in the medically ill: a comparison of self-report measures, clinician judgment, and DSM-IV diagnoses. *Psychosomatics* **45**, 461–469.
- Williams JW, Noel PH, Cordes JA, Ramirez G, Pignone M** (2002a). Is this patient clinically depressed? *Journal of the American Medical Association* **287**, 1160–1170.
- Williams JW, Pignone M, Ramirez G, Stellato CP** (2002b). Identifying depression in primary care: a literature synthesis of case-finding instruments. *General Hospital Psychiatry* **24**, 225–237.
- Youden WJ** (1950). Index for rating diagnostic tests. *Cancer* **3**, 32–35.
- Zimmerman M, McGlinchey JB, Young D, Chelminski I** (2006a). Diagnosing major depressive disorder IV: relationship between number of symptoms and the diagnosis of disorder. *Journal of Nervous and Mental Disease* **194**, 450–453.
- Zimmerman M, McGlinchey JB, Young D, Chelminski I** (2006b). Diagnosing major depressive disorder I: a psychometric evaluation of the DSM-IV symptom criteria. *Journal of Nervous and Mental Disease* **194**, 158–163.
- Zimmerman M, McGlinchey JB, Young D, Chelminski I** (2006c). Diagnosing major depressive disorder IX: are patients who deny low mood a distinct subgroup? *Journal of Nervous and Mental Disease* **194**, 864–869.