Despite being common, depression often goes undetected in patients with skin diseases. Our aim was to examine and compare the performance of three depression screeners. We studied dermatological inpatients aged 18–65 years. They completed the questionnaires Primary Care Screener for Affective Disorders (PC-SAD), Patient Health Questionnaire (PHQ) and General Health Questionnaire (GHQ-12) and were administered a standardized psychiatric interview (SCID-I) by a mental health professional, who was unaware of the questionnaire answers. The analysis was performed on 141 patients with complete data (79% of all eligible patients, 89% of all patients who agreed to participate). The prevalence of the main forms of depression, major depressive disorder and dysthymic disorder, was 8.4% and 6.3%, respectively. For major depressive disorder, the sensitivity and specificity of the questionnaires were as follows: PC-SAD, 73% and 88%; PHQ, 55% and 91%; GHQ-12, 73% and 78%. For dysthymic disorder, the sensitivity and specificity were as follows: PC-SAD, 56% and 95%; PHQ, 44% and 90%; GHQ-12, 56% and 76%. The small sample size suggests caution in drawing conclusions about the relative merits of these screeners. Although both the GHQ and the PHQ are short and easily hand scored, the first is a generic screener for psychiatric morbidity that is not specific for depression, while the second displayed modest sensitivity. The PC-SAD, with short average administration time, acceptable sensitivity and high specificity, might be particularly useful in settings where the technology for computer automated scoring is available. Although screening programmes might be useful, they should be supplemented by quality improvement programmes and by the development of consultation-liaison services. Key words: depression; diagnosis; skin disease; public health.

(Accepted February 7, 2005.)


Angelo Picardi, Dermatological Institute IDI-IRCCS, Via dei Monti di Creta, 104, IT-00167 Rome, Italy. E-mail: a.picardi@idi.it

Epidemiological studies have shown that depressive disorders are common, morbid and treatable (1). Given their prevalence and consequences in terms of excessive mortality, disability and secondary morbidity, depressive disorders represent a major public health problem (2). The World Health Organization Global Burden of Disease Survey has estimated that by the year 2020, major depression will be second only to ischaemic heart disease in the amount of disability experienced by sufferers (3).

In non-psychiatric medical settings, depressive disorders are even more frequent than in the general population (4). Among patients with skin diseases, psychiatric disorders are frequent (5-8) and depressive disorders, in particular, were found in almost 10% of patients (6), while in another study 4% of patients expressed active suicidal ideation and 7.3% endorsed a wish to be dead (9). Co-morbid depression can exacerbate the effects of medical illness and may be an independent source of suffering and disability (10). Psychiatric morbidity is also associated with poor medication adherence (11).

As effective treatments for depression are available, timely and appropriate interventions might relieve the burden of depression on individuals, health services and society (2). However, in a high proportion of patients depression is neither recognized nor treated adequately (12). Some studies have documented that under-recognition of depression is also frequent in dermatology outpatient clinics (5, 13, 14). Therefore, improving diagnostic efficiency for depressive disorders in dermatology clinics, primary care and other non-psychiatric settings continues to be an important first step in addressing a major public health issue (2).

Although many depression screening tools have been developed, not all are suitable for routine use (15, 16). Lengthy self-administered and clinician-administered instruments are impractical. Instruments that do not adhere to established diagnostic criteria may not indicate which clinical symptoms of depression are present, make physician follow-up more difficult and time-consuming, and require a greater deal of trust in the validity of the instrument.

In the present study, we used a standardized psychiatric interview to determine the presence of depression in a sample of dermatological inpatients,
and we tested the screening properties of a widely used generic screener for non-psychotic psychiatric disorders and of two depression screeners meeting the requirements of brevity, self-completion and diagnostic transparency.

**MATERIALS AND METHODS**

**Participants and procedure**

The study was carried out at the inpatient wards of the IDI-IRCCS, a large dermatological hospital located in Rome, Italy. This institution serves the entire population of Rome and its province (approximately 3,800,000 people, of whom 320,000 are aged 18 or over). Also, some patients are referred to IDI-IRCCS from other regions, mainly from central and southern Italy.

The institutional review board approved the study protocol. All patients aged ≥18 years, free from dementia or severe cognitive impairment, admitted to five of the eight inpatient wards of IDI-IRCCS on predetermined days were contacted by a research dermatologist. All patients who agreed to participate gave written informed consent. The research dermatologist collected demographic and clinical information and gave participants a research questionnaire to complete. This questionnaire included several instruments, including the 12-item General Health Questionnaire (GHQ-12), the Patient Health Questionnaire (PHQ) and the Primary Care Screener for Affective Disorders (PC-SAD), in that order at each administration. Within 48 h, all participants were also administered the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) by a trained mental health professional, who was unaware of questionnaire answers.

A total of 178 eligible patients were contacted by the research dermatologist. Of these, 139 (89%) accepted, signed the informed consent form and were given the research questionnaire and scheduled to be interviewed within 48 h. Nine patients were discharged before the scheduled interview. Of the 150 patients who were administered the SCID-I, five returned an incomplete PHQ, three returned an unusable PC-SAD, and one was excluded due to missing items on the GHQ-12. Therefore, the study sample consists of 141 patients with valid data for all instruments (79% of all eligible patients, 89% of all patients who agreed to participate).

**Instruments**

The SCID-I (17) is a standardized psychiatric interview yielding psychiatric diagnoses according to the 4th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (18). We used the official Italian version (19). Interrater reliability was tested with all interviewers on two separate occasions, and the overall kappa coefficient was found to be constantly above 0.80.

The PHQ (20) is a self-administered version of the PRIME-MD (Primary Care Evaluation of Mental Disorders) (21). Its diagnostic validity is corroborated by the strong association of PHQ diagnoses with indices of functional impairment and health care services use. It has also been tested against independent diagnoses made by mental health professionals (20). We used the nine-item section on depressive disorders of the Italian version (kindly provided by Pfizer, Italy).

The PC-SAD (16) is a recently developed short, self-administered questionnaire, constructed in accordance with DSM-IV criteria. Its validity has undergone preliminary testing against other established screening questionnaires (16), and the instrument has been used recently to identify depressed patients to be included in a randomized clinical trial (22).

An important feature of the PC-SAD is that it allows clinicians to screen not only for major depressive disorder (MDD) but also for dysthymic disorder. While MDD is characterized by depressed mood or markedly diminished interest or pleasure in daily activities lasting at least 2 weeks plus at least four additional depressive symptoms, dysthymic disorder is a chronic form of depression characterized by mild to moderate symptoms lasting at least 2 years that often goes unrecognized and untreated in spite of its frequency, its impact on general health and quality of life, and the availability of effective treatments (23).

The PC-SAD consists of a 3-item prescreener, a 26-item section and an 8-item dysthymic disorder section. The prescreener questions are part of the screener score, but reduce respondent burden by terminating the questionnaire if all are negative. The average number of completed items is about 10, with an average time of completion of less than 2 minutes (16).

The PC-SAD differs from the PHQ because it breaks down each of the nine DSM-IV symptoms of MDD into several simple items, each consisting of single concept questions, and it integrates the answers mathematically. As compared with the PHQ, the PC-SAD questions are intentionally shorter and use simpler language. The responses related to a given symptom are numbered and summed, and if the sum is above a certain threshold, the patient is said to have the symptom. The questionnaire is scored using an automated system. The scoring algorithm is built in a way that the presence of each symptom can be determined independently from the presence of missing answers to one or more items related to the symptom, provided that at least one of the items related to the symptom has been answered. Hence, the questionnaire can give valid results even if many items are left unanswered, provided that at least one item for each DSM-IV symptom is answered. In practice, a questionnaire with 50% or more completed items is usually valid, whereas even questionnaires with a lower proportion of completed items can be valid, depending on the specific profile of responses in each case. In this study, only three PC-SAD questionnaires were unusable because of missing answers.

To obtain a valid Italian version of the PC-SAD, we followed guidelines for the cross-cultural adaptation of health measures (24). An initial translation was produced by another psychiatrist (AP) with previous experience in translating English instruments into Italian. This first version was independently revised by a psychiatrist (PP), a clinical psychologist, and an epidemiologist (DA). All suggestions were discussed by the translator with the reviewers, and those deemed to be relevant were included in a second version. This process was repeated one more time, and the third version was back-translated by a bilingual colleague. The first back-translation was thoroughly reviewed by three authors of the original questionnaire (DA, WHR and KMB). A second back-translation was produced and reviewed again, until consensus was reached.

The GHQ-12 is a self-administered questionnaire designed to measure psychological distress and to detect general psychiatric morbidity in medical settings and in the community (25). The Italian version has been shown to be valid and reliable in patients with skin diseases (26). At variance with the PHQ and the PC-SAD, it does not produce a DSM-IV diagnosis, but identifies a patient as a ‘probable case of psychiatric disorder’. It is scored by summing up the scores on each item, rather than using a scoring algorithm.
Data reduction and statistical analysis

The PHQ scoring algorithm was used to generate the diagnoses of MDD and ‘other depressive disorder’. The PC-SAD scoring algorithm was used to generate the diagnoses of MDD and dysthymic disorder. The GHQ-12 was scored with the binary scoring method, collapsing adjacent responses to obtain a dichotomous scoring (0-0-1-1).

Using the SCID-I as the criterion standard, we computed sensitivity, specificity, and predictive values of the three questionnaires. Sensitivity is defined as the proportion of people with a disease who have a positive test result, whereas specificity is the proportion of people without the disease who have a negative test result. The positive predictive value (PPV) is the proportion of patients with a positive test result who have the disease, while the negative predictive value (NPV) is the proportion of patients with a negative test result who are free from the disease. Unlike predictive values, sensitivity and specificity are largely independent of the setting in which a test is used because they are quite stable with changes in prevalence of the disease in the population studied (27).

All three instruments were compared in their ability to detect MDD. To identify the best GHQ-12 cut-off score for MDD, we performed receiver operating characteristic (ROC) analysis (28). The GHQ-12 cut-off score with the best balance between sensitivity and specificity was identified and used in the analyses.

The questionnaires were also compared in their ability to detect dysthymic disorder. Given that the PHQ does not have a scoring algorithm specific for this disorder, we studied the performance of its generic scoring algorithm for ‘Other depressive disorders’. The ROC analysis was used to identify the best GHQ-12 cut-off score for dysthymic disorder.

Exact binomial 95% confidence intervals for proportions were computed with the Epi-Info software (29). All other analyses were run under SPSS, version 8.0 for Windows.

RESULTS

The mean age of participants was 37.5 years and 44% were men. Their sociodemographic and clinical characteristics are summarized in Table I.

Forty-four participants received a DSM-IV diagnosis of some depressive condition on the SCID-I. Twelve satisfied the DSM-IV criteria for MDD (8.4%), although one was in partial remission and currently did not meet the criteria for a depressive episode. Nine patients satisfied the criteria for dysthymic disorder (6.3%). Of these, one had ‘double depression’ (concurrent MDD and dysthymic disorder). One patient had a diagnosis of bipolar I disorder, depressive phase. The other 23 patients had broadly defined depressive conditions, such as depressive disorder not otherwise specified (7.1%), adjustment disorder with depressed mood (7.1%), or adjustment disorder with anxiety and depressed mood (2.1%). While the first condition is a form of depression that, although clinically significant, does not meet the full criteria for severity, duration or level of impairment of MDD or dysthymic disorder, the latter two conditions are relatively mild forms of depression in which depressive symptoms develop in response to a stressful event or situation.

The performance of the three instruments in detecting the 11 patients who currently met the criteria for MDD by the SCID is detailed in Table II.

The PC-SAD identified 24 patients as cases of MDD, of whom 8 had MDD according to the SCID, and 16 did not. Hence, its sensitivity was 73% because it detected 8 of the 11 patients with MDD as determined by the SCID, while its specificity was 88% because it correctly classified as free from MDD 114 of 130 patients who did not meet the criteria for MDD by the SCID.

The PHQ identified 18 patients as positive for MDD, of whom 6 had MDD according to the SCID, and 12 did not (sensitivity 55%, specificity 91%).

As regards the GHQ-12, we found that a score of 7 or higher best identified the cases of MDD. This cut-off score is much higher than the usual cut-off scores in

Table I. Sociodemographic and clinical characteristics of participants

<table>
<thead>
<tr>
<th>Parameter</th>
<th>n  (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>62 (44.0)</td>
</tr>
<tr>
<td>Female</td>
<td>79 (56.0)</td>
</tr>
<tr>
<td>Age range (years)</td>
<td></td>
</tr>
<tr>
<td>18–29</td>
<td>54 (38.3)</td>
</tr>
<tr>
<td>30–39</td>
<td>28 (19.9)</td>
</tr>
<tr>
<td>40–49</td>
<td>27 (19.1)</td>
</tr>
<tr>
<td>50–59</td>
<td>22 (15.6)</td>
</tr>
<tr>
<td>60 or more</td>
<td>10 (7.1)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
</tr>
<tr>
<td>Unmarried</td>
<td>57 (40.4)</td>
</tr>
<tr>
<td>Married</td>
<td>80 (56.7)</td>
</tr>
<tr>
<td>Separated, divorced, widower, or widow</td>
<td>4 (2.8)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
</tr>
<tr>
<td>&lt;12th grade</td>
<td>39 (27.7)</td>
</tr>
<tr>
<td>High school graduate</td>
<td>81 (57.4)</td>
</tr>
<tr>
<td>College graduate</td>
<td>21 (14.9)</td>
</tr>
<tr>
<td>Dermatological diagnosis</td>
<td></td>
</tr>
<tr>
<td>Psoriasis</td>
<td>51 (36.2)</td>
</tr>
<tr>
<td>Bullous disease</td>
<td>4 (2.8)</td>
</tr>
<tr>
<td>Alopecia</td>
<td>1 (0.7)</td>
</tr>
<tr>
<td>Acne</td>
<td>2 (1.4)</td>
</tr>
<tr>
<td>Skin tumours</td>
<td>4 (2.8)</td>
</tr>
<tr>
<td>Vitiligo</td>
<td>2 (1.4)</td>
</tr>
<tr>
<td>Urticaria</td>
<td>7 (5.0)</td>
</tr>
<tr>
<td>Contact dermatitis</td>
<td>2 (1.4)</td>
</tr>
<tr>
<td>Atopic dermatitis</td>
<td>6 (4.3)</td>
</tr>
<tr>
<td>Other forms of dermatitis</td>
<td>20 (14.2)</td>
</tr>
<tr>
<td>Connective tissue disease</td>
<td>5 (3.5)</td>
</tr>
<tr>
<td>Cutaneous vasculitis</td>
<td>3 (2.1)</td>
</tr>
<tr>
<td>Skin ulcers</td>
<td>4 (2.8)</td>
</tr>
<tr>
<td>Bacterial infections</td>
<td>5 (3.5)</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>16 (11.3)</td>
</tr>
<tr>
<td>Missing information or diagnosis not definitively established</td>
<td>9 (6.4)</td>
</tr>
</tbody>
</table>

Missing information or diagnosis not definitively established

| Duration of skin disease (years)               |        |
| 0–3                                           | 70 (49.6) |
| 4–9                                           | 18 (12.8) |
| 10–15                                         | 20 (14.2) |
| >15                                           | 20 (14.2) |
| Missing information                           | 13 (9.2) |
studies screening for general psychiatric morbidity, e.g. including also anxiety, somatoform and adjustment disorders. With this high cut-off score, 37 patients scored positive, of whom 8 had MDD according to the SCID, and 29 did not (sensitivity 73%, specificity 78%).

For MDD, the PPV of the PC-SAD, the PHQ and the GHQ in our patient population was 33%, 33%, and 22%, respectively.

Table II also reports the performance of the three instruments in detecting the nine cases of dysthymic disorder as diagnosed by the SCID.

The PC-SAD identified 12 patients as cases of dysthymic disorder, of whom 5 were SCID-positive (sensitivity 56%, specificity 95%). Of the seven patients who scored positive without having a SCID diagnosis of dysthymic disorder, all except one was affected by another depressive condition: three had a depressive disorder not otherwise specified, one had recurrent MDD and two had a single episode MDD.

The PHQ identified 17 patients as positive for a depressive disorder other than MDD. Among them, there were only three of the nine patients with dysthymic disorder. A fourth patient, who had double depression and was identified by the PHQ as having MDD, was considered as correctly classified, because in the PHQ scoring algorithm the diagnosis of MDD includes and rules out the diagnosis of ‘other depressive disorder’. Hence, sensitivity was 44% and specificity was 89%.

Also for dysthymic disorder, the best GHQ-12 cut-off score was found at 6/7. Of the 37 patients who scored above this threshold, 5 had dysthymic disorder (sensitivity 56%, specificity 76%).

For dysthymic disorder, the PPV of the PC-SAD, the PHQ and the GHQ in our patient population was 42%, 24% and 14%, respectively.

DISCUSSION

This study attempted to delineate the relative merits and limitations of three brief depression screeners that are suitable for routine use in busy clinical practices.

The PC-SAD displayed good sensitivity and excellent specificity for MDD. As regards dysthymic disorder, sensitivity was low, possibly as a result of patients’ difficulties in answering questions about their mood state over a long time period. Conversely, the specificity was very high. Also, given that most false positives were actually affected by another depressive condition, a positive result for dysthymia suggests closer attention to the probable presence of depression.

As in other studies (20, 30), the PHQ displayed excellent specificity for MDD. Specificity was higher than the average value reported for other screeners (15, 31). However, its sensitivity was barely acceptable. Similarly, high specificity but modest sensitivity were found for dysthymic disorder, for which the PHQ does not provide a definite diagnosis, but only an indication that a depressive disorder other than MDD is present. Although the PHQ performed better in the original validation study (20) than in ours, some differences between studies prevent direct comparison. The validation study was partly based on the original PRIME-MD, and was carried out over the telephone. Also, in our study the PHQ was fully self-administered, whereas in the original study the general practitioner briefly reviewed the questionnaire with each patient and asked any additional questions necessary to clarify responses to the questionnaire.

As regards the GHQ-12, for both MDD and dysthymic disorder the best balance between sensitivity and specificity was obtained with a very high cut-off score of 6/7. Using this cut-off, sensitivity and specificity for MDD were similar to the average values reported for dysthymic disorder, of whom 5 were SCID-positive (sensitivity 56%, specificity 95%). Of the seven patients who scored positive without having a SCID diagnosis of dysthymic disorder, all except one was affected by another depressive condition: three had a depressive disorder not otherwise specified, one had recurrent MDD and two had a single episode MDD.

The PC-SAD identified 12 patients as cases of dysthymic disorder, of whom 5 were SCID-positive (sensitivity 56%, specificity 95%). Of the seven patients who scored positive without having a SCID diagnosis of dysthymic disorder, all except one was affected by another depressive condition: three had a depressive disorder not otherwise specified, one had recurrent MDD and two had a single episode MDD.

The PHQ identified 17 patients as positive for a depressive disorder other than MDD. Among them, there were only three of the nine patients with dysthymic disorder. A fourth patient, who had double depression and was identified by the PHQ as having MDD, was considered as correctly classified, because in the PHQ scoring algorithm the diagnosis of MDD includes and rules out the diagnosis of ‘other depressive disorder’. Hence, sensitivity was 44% and specificity was 89%.

Also for dysthymic disorder, the best GHQ-12 cut-off score was found at 6/7. Of the 37 patients who scored above this threshold, 5 had dysthymic disorder (sensitivity 56%, specificity 76%).

For dysthymic disorder, the PPV of the PC-SAD, the PHQ and the GHQ in our patient population was 42%, 24% and 14%, respectively.
other screeners (15, 31), whereas they were barely satisfactory for dystymic disorder.

The PPV was low for all three questionnaires. Indeed, at the prevalence rates of depression usually found in dermatological or primary care settings, all screeners have low PPV. For this reason, some experts suggest limiting the screening to patients known to be at higher risk, such as those with personal or family history of depression, unexplained physical symptoms, chronic pain, or higher-than-expected use of medical services (32, 33). In this way, the screened population would be 'selectively enriched' with depressed patients and the PPV would substantially increase. For instance, doubling the prevalence rate of MDD in our sample would raise to 50% the PPV of the PC-SAD, whereas trebling the prevalence rate would bring it to 65%.

This study has some limitations. First, the relatively small sample size prevents firm conclusions being drawn about the relative merits of the screeners tested.

Second, our findings might not generalize to other medical settings. However, our sample should resemble in many respects patients who are seen by general practitioners or other non-psychiatry specialists. Indeed, the prevalence rates of MDD and dystymic disorder in our sample were similar to those observed in primary care patients (4).

Third, we studied only inpatients. However, many patients of mild or moderate severity coming from more disadvantaged Italian areas are admitted as inpatients to IDI-IRCCS for diagnostic assessments and treatments not easily available in their regions. Hence, our results could probably be generalized to outpatients in other health systems where hospitalization for skin diseases is rarer.

The fourth limitation is the time lapse between the administration of the screeners and the interview. However, the SCID-I was administered within 48 h, usually after 24 h. Hence, substantial changes in clinical state are unlikely to have occurred.

Fifth, the questionnaires were always administered in the same order, raising the possibility of bias due to order of completion. This possible bias, however small, could have been prevented with a randomized scheme of completion.

Finally, it should be noted that different authors have translated the assessment instruments into Italian. Except the GHQ-12, all instruments enquire about each symptom of MDD as defined in the original English language versions. Although there are no notable problems or discrepancies in any of the Italian versions of the instruments, some subtle differences in the meaning or definitions of one or more symptoms might have slightly affected the results.

This study confirmed that depression is frequent among dermatological patients. Depression causes substantial suffering and disability, and might impair adherence to dermatological treatment (11). The importance of co-morbid depression in dermatological patients is further underscored by reports of suicidal ideation (9) and completed suicide (34). Unfortunately, depression often seems to go undetected in dermatological patients (13, 14). Given that it can be effectively treated, efforts should be devoted to increasing its recognition. The use of a screening questionnaire may help dermatologists to identify depressed patients and to provide either direct treatment or referral to a mental health professional.

The three questionnaires studied represent reasonable alternatives to screen for depression. The GHQ-12 is short, easy to score, and has acceptable sensitivity and specificity with the use of an unusually high cut-off score. However, a positive result requires much further probing from the clinician because the instrument provides neither a tentative screener diagnosis nor a list of individual DSM-IV symptoms of depression. Hence, the GHQ is probably best used to screen for general psychiatric morbidity rather than specifically depression. The PHQ is short, easily scored, and has high specificity despite modest sensitivity. It might be useful in settings where missing some cases of depression is not considered a major problem, a quick diagnosis is needed, and computer-scoring methods are not available. In settings where the technology for computer automated scoring is available, the PC-SAD, with short average administration time, acceptable sensitivity and high specificity, might be the instrument of choice.

Although simple and practical depression screeners are helpful, it should be emphasized that screening and increased recognition alone are insufficient, particularly in settings lacking systems to assure accurate diagnosis, effective treatment and careful follow-up (20, 32). Screening programmes should be supplemented by quality improvement programmes and by the development of consultation-liaison services enabling a fruitful, mutual collaboration between dermatologists and mental health professionals (35).

ACKNOWLEDGEMENTS

This study was financially supported by the Italian Ministry of Health in the framework of the “Progetto Ricerca Corrente”. We are grateful to Pfizer Italia S.r.l. for kindly providing us with the Italian version of the PHQ and for permission to use it. We thank Elisabetta Luchetti, Nidia Melo Salcedo and Valentina Salvatori for their patient help in data collection, data input and quality control. We are indebted to Daniela Bitetti for her invaluable help in the process of translating the PC-SAD into Italian, and to Simone Bolli for his precious technical assistance in data management.

REFERENCES

1. Horwath E, Weissman MM. Epidemiology of depression and anxiety disorders. In: Tsuang MT, Tohen M,