Suffering in silence: studies on screening for major depressive disorder in primary care
Wittkampf, K.A.

Citation for published version (APA):
The psychometric properties of the panic disorder module of the Patient Health Questionnaire (PHQ-PD) in high-risk groups in primary care


Abstract

Aims
To study the validity of detecting panic disorder (PD) using the Patient Health Questionnaire (PHQ) in a high-risk population in primary care and to test whether modified evaluation algorithms improve the operating characteristics of this questionnaire. In addition, the influence of psychiatric comorbidity on the test characteristics of the panic module was studied.

Methods
The PHQ was administered in a primary care sample with patients at high-risk for psychiatric disorders. The total sample of 479 high-risk patients comprised 311 frequent attenders (FA), 39 patients with unexplained somatic complaints (USC) and 191 patients with mental health problems (MHP). The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID- I) was the reference standard for the presence of PD. Sensitivity, specificity, and predictive values were calculated. The conditional test characteristics were calculated based on the observed prevalence of PD in the three high-risk groups.

Results
PD was diagnosed in 4.8% of the FAs, in 9.8% of the USC and in 7.6% of the MHPs. The PHQ achieved moderate operating characteristics. Modified evaluation algorithms of the questionnaire led to an improvement of test characteristics, especially the screening question: sensitivity .71 and specificity .83. Psychiatric comorbidity increased sensitivity while decreasing specificity.

Conclusion
The original and modified algorithms of the PHQ-PD performed moderately in screening for panic disorder. Using only the first question of the PHQ-PD showed the best psychometric properties (sensitivity). For screening purposes requiring high sensitivity we endorse to use the screening question instead of the original algorithm.
Introduction

Panic disorder affects 2% to 3% of the adult population and the median prevalence of panic disorder in primary care setting is 4–6%. Former research suggests that panic disorder often goes unrecognised in primary care. This seems unacceptable because patients with panic disorder experience considerable impairment and disability, including occupational difficulties, impaired wellbeing and reduced quality of life. They also show increased suicidal ideation and suicidal attempts, even after adjustment for affective comorbidity and other suicide risk factors. More importantly, response to treatment is good with a 70% recovery rate.

Methods to improve the recognition of panic disorder in primary care include the application of short screening questionnaires. So far screening for psychiatric disorders in primary care attendees seems not effective because screening instruments have low positive predictive values in unselected populations, related to the relatively low prevalence of (undisclosed) disorders. Hence, screening yields many false-positive results. Screening in high-risk groups seems to be a reasonable alternative in reducing this number of false positives.

The Patient Health Questionnaire (PHQ), a short self-reporting questionnaire derived from the PRIME-MD interview, is designed to screen for psychiatric disorders according to the DSM-IV criteria. Currently, the PHQ is used all over the world and has been translated into more than 25 languages, including German, French, Spanish, Italian, Arabic, Bengali, Turkish, Flemish and Dutch. Besides major depressive disorder, other anxiety disorder, somatoform disorder, alcohol abuse and eating disorders, the PHQ is designed to screen for panic disorder (PHQ-PD). The test characteristics of the depression module of the PHQ have been widely studied, but so far little information is available about the test characteristics of PHQ-PD used in primary care populations. The evidence that is available varies widely, especially sensitivity, because of the relatively low prevalence of the disorder; former reported sensitivities varied from 0.17 to 0.57. Furthermore one study has reported that using modified scoring algorithms or just one screening question instead of the original algorithm for the PHQ-PD could improve the screening performance of the test.

Screening within a high-risk population in primary care for PD with the PHQ-PD has not been studied before. Furthermore it is unknown whether a positive score on other PHQ-modules influences the psychometric properties of the PHQ in diagnosing PD. This might be important for two reasons. First, because patients suffering from panic disorder might as well suffer from depressive, anxiety and/or somatoform disorders. Secondly, some symptoms are included in the definition of both panic and somatoform disorder (for instance feeling your heart pounce,
shortness of breath and dizziness). Furthermore many patients with depressive symptoms or somatoform symptoms are also suffering from anxiety or panic but not because they have a comorbid anxiety disorder, but because anxiety and panic symptoms can be part of the depressive or somatoform disorder. Therefore patients with depressive or somatoform symptoms might have higher scores on the PHQ-PD compared to patients without any psychiatric symptoms, thereby decreasing specificity and increasing sensitivity.

We aimed to study the test characteristics of the PHQ-PD in a high-risk population in primary care. We compared the PHQ diagnoses to the SCID-I, a widely used reference standard for the diagnosis of psychiatric disorders. Our secondary aim was to test whether the screening question or modifications in the questionnaires' evaluation algorithms lead to better operating characteristics. Finally we aimed to study the influence of scores on other modules of the PHQ (PHQ-depression and PHQ-somatisation) on the test characteristics of the PHQ.

**Methods**

This study was part of a larger study in which the complete PHQ was validated. Methodological details and results of the validation of the depression module and somatoform module were published elsewhere.\(^{21,22}\) We studied screening and diagnostic test characteristics of the PHQ-PD within a high-risk population for psychiatric disorders in primary care. We considered the first question of the PHQ-PD module as a screener and we studied the test characteristics of this screener as well as of the original algorithm that applies DSM-IV diagnostic criteria.\(^{17}\)

**Study population**

The study population has been described in detail in our former validation study of the PHQ-mood module.\(^{21}\) In summary three groups of patients at risk for psychiatric disorders were selected.

1. Patients with unexplained somatic complaints (USC). These complaints cannot be explained in medical terms according to the general practitioner (GP). These complaints had to be present for at least three months. GPs selected these patients from consultation lists during four weeks.

2. Patients with consultation rates in the highest 10% (frequent attenders, FA) in the year preceding study allocation. We used the method as proposed by Howe.\(^{23}\) The 10% most frequently consulting women and the 10% most frequently attending men in 2 age-groups (18 to 44 and 45 to 70 years) in the year preceding study allocation. This method accounted for differences in sex and age among frequently attending patients. We determined the highest 10% consulting patients separately for each GP to account for differences in practice styles.
3. Patients presenting to their GP with a new mental health problem up to three months prior to the selection date (Mental Health Problem, MHP). We selected these patients from electronic patient databases of the participating GP practices. Three months was chosen as a timeframe because of the transitory nature of most mental health problems.

**Procedures**

GPs received a list of selected patients and they excluded those patients who were suffering from schizophrenia, psychosis or bipolar disorder, serious somatic disease, or mental retardation, or having difficulties with Dutch or English language. Subsequently, the PHQ was sent to the patients and if patients did not respond within two weeks a reminder was sent. Missing values on the PHQ were solved by means of telephone contact.

**Sampling procedure**

We used a two-stage design in which all participants completed the PHQ screening questionnaire and were rated as cases at risk for depressive disorder or cases not at risk for depressive disorder in accordance with standardized criteria. In the second stage, we invited all cases and a random selection of 20% of non-cases to fill in a second PHQ just before completion of the reference standard, a SCID-I interview. This second PHQ was used for the present analysis because the time interval between the screener and reference standard had to be limited (max 1 week). All prevalence figures have been corrected for this random selection with inversed probability weighing.

**Measurement instruments**

*Patient Health Questionnaire-panic disorder*

The PHQ is developed to make a criterion-based diagnosis of mental disorders commonly encountered in primary care. The panic module is one part of the PHQ (questions 3a–d and 4a–k). Question 3 includes items derived from the DSM-IV classification system to ask for the history and frequency of panic attack: (3a) In the last 4 weeks, have you had an anxiety attack — suddenly feeling fear or panic? (3b) Has this ever happened before? (3c) Do some of these attacks come suddenly/out of the blue — that is, in situations where you don’t expect to be nervous or uncomfortable? (3d) Do these attacks bother you a lot or are you worried about having another attack? Question 3a is a screening question. If answered ‘no’, subjects are asked to proceed to the next module. Question 4 includes the somatic symptoms of panic attack: (4a) shortness of breath, (4b) heart palpitation, (4c) chest pain or chest pressure, (4d) sweating, (4e) feelings of choking, (4f) hot flashes or chills, (4g) nausea or an upset stomach, or diarrhea, (4h) feeling dizzy, unsteady, or faint, (4i) tingling or numbness in parts of your
body; (4j) trembling or shaking; and (4k) afraid of dying. There are two answer categories: ‘no’ (0 points) and ‘yes’ (1 point).

A positive score on the panic module is indicated if questions 3a–d are all answered with yes (sum score 4) in combination with four or more other items of question 4 answered with yes (sum score 4). However, since a high detection rate is the first objective for screening instruments, we tested whether modified evaluation algorithms would increase the sensitivity for diagnosing panic disorder. The original evaluation algorithm requires that the first four responses to question 3 are positive. The modified algorithms (i and ii) require two (or three) positive responses, of which 3a.

The single screening question for which we tested diagnostic validity was taken from the PHQ panic module (question 3a). The exact wording is: ‘In the last 4 weeks, have you had an anxiety attack - suddenly feeling fear or panic?’

*Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I)*

The SCID-I is a semi-structured interview for diagnosing mental disorders on Axis-I according to DSM-IV criteria. The SCID-I was administered by researchers after training by a skilled professional. Throughout the study all interviewers had continuing training sessions supervised by an expert psychiatrist. They also had monthly consensus meetings to maximise accuracy and consistency in the administration of the SCID-I.

This SCID-I interview was administered by telephone. The interviewers were blinded for the PHQ results. Agreement between diagnosis gained from telephone and live administration of the SCID-I has been found to be excellent (Kappa 0.73 (with 90% agreement)).

**Statistical analysis**

*Prevalence*

For the calculation of the prevalence of panic disorder in our population, we had to correct for the fact that we took a random sample of all patients with a low risk on PD (see sampling procedure). We corrected this imbalance with inverse probability weighting. All calculations have been performed with these corrected data because otherwise the disease prevalence and test characteristics (sensitivity and specificity) will be different as a result of our sampling procedure.
**Criterion validity**

We assessed the criterion validity of the second PHQ by calculating the test characteristics (sensitivity, specificity, positive and negative predictive value) using several scoring criteria, including the original algorithmic criterion, two modified algorithms (i and ii) and the screening question alone. Ninety-five per cent confidence intervals for proportions were calculated according to the efficient-score method (corrected for continuity). Additionally, differences of sensitivities and specificities of the original algorithm compared to the two modified algorithms and to the screening question were tested using McNemar X2 tests with the Bonferonni–Holm procedure to adjust for multiplicity. Missing values on the panic disorder question (n = 42, 8.8%) on the second PHQ were solved by imputing the answers of the first PHQ. We studied the influence of a positive score on the depression module and of the somatisation module of the PHQ, but not of other PHQ-modules (other anxiety disorder, eating disorder and alcohol problems) because these modules are not well validated yet.

**Results**

**Study population**

Six health centers with 23 GPs and 31915 enlisted patients participated. A total of 2,659 patients (8.3%) fulfilled the criteria for frequent attending (FA; 1745), mental health problems (MHP; 1049) and/or unexplained somatic complaints (USC; 183). GPs excluded 345 (13.0%) patients for screening. Of 2314 patients eligible for screening with the Patient Health Questionnaire (PHQ), 1029 (44.5%) patients returned the questionnaire; 980 (42.4%) patients gave informed consent. When comparing patients that gave informed consent (n = 980) to patients that did not respond or did not give informed consent (n = 1,334, non-responders), the first group was older (mean age 49.7 versus 44.2 years, p ≤ 0.05) and consisted of more females (63.7% versus 58.7%, p ≤ 0.05). Furthermore this group consisted of less MHP- patients compared to the non-responders although this difference was small (38.2% versus 42.5%, p ≤ 0.05). No significant differences were found considering distribution of the other two high-risk groups (FA: 65.6% versus 62.3%; USC: 7.1% versus 7.4%, ≤ 0.05).

From the patients with above threshold scores (n = 579) we reached 418 patients (72%) and from the selected sample of patients with below threshold scores (n = 82) we reached 61 patients (74%) for the second PHQ (the one used for analysis) and SCID-I (total n = 479). When comparing patients that were included for final analysis (n = 479) to patients that could not be reached (n=180), no significant differences were found considering age, sex ratio and distribution among high-risk groups. Sixty-one of 479 patients belonged to the random sample of 20% (see
Methods section); to correct for this sampling, these 61 patients were weighted with a factor of 5 \((61 \times 5 = 305)\) (inversed probability weighing).\textsuperscript{25} All following calculations were performed with the corrected population \((n = 305 + 418 = 723)\) (see flowchart in Fig.1 and Venn diagram in Fig. 2).

**Figure 1.** Flowchart of patients.

![Flowchart of patients](image-url)

- Total selection 2659
  - Exclusion 345
    - Psychotic/ bipolar disorder (43)
    - Language problems (49)
    - Somatic problem/ mental retardation (71)
    - Other (objection GP, deceased) (182)
  - Invited 2314
    - No response 1285
    - No informed consent 49
    - Response 1029 (44%)
    - Informed consent 980
      - PHQ+ (depression, anxiety or somatoform disorder) 579
        - Reached for SCID 418 (72%)
          - Analysis 418
          - Panic disorder 41 (5.7%)
        - Missing 1
          - PHQ- (depression, anxiety or somatoform disorder) 400
            - Reached for SCID 61 (74%)
              - Analysis 305
                - IPW* factor 5
  - Total 723

*IPW: inversed probability weighting
**Figure 2.** Venn diagram of high-risk groups: frequent attenders (FA), patients with mental health problems (MHP) and patients with unexplained somatic complaints (USC). Number of patients, between brackets: patients with panic disorder.

![Venn diagram](image)

**Table 1.** Study population (N= 723)

<table>
<thead>
<tr>
<th>Population</th>
<th>Patients [n (%)]</th>
<th>Females [n (%)]</th>
<th>Mean age (years)</th>
<th>Panic disorder [n (%)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study population</td>
<td>723 (100)</td>
<td>477 (66)</td>
<td>48.9</td>
<td>41 (5.7)</td>
</tr>
<tr>
<td>FA*</td>
<td>495 (68.5)</td>
<td>298 (60.2)</td>
<td>49.1</td>
<td>24 (4.8)</td>
</tr>
<tr>
<td>MHP*</td>
<td>251 (34.7)</td>
<td>179 (71.3)</td>
<td>47.3</td>
<td>19 (7.6)</td>
</tr>
<tr>
<td>USC*</td>
<td>51 (7.1)</td>
<td>47 (92.2)</td>
<td>50.7</td>
<td>5 (9.8)</td>
</tr>
</tbody>
</table>

FA frequent attenders, MHP patients with mental health problems, USC patients with unexplained somatic complaints. *A considerable number of patients belonged to more than one high-risk group.
### Table 2. Test characteristics of PHQ-panic disorder (PD).

<table>
<thead>
<tr>
<th>Cut off point</th>
<th>Total</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>Sensitivity</th>
<th>McNemar $\chi^2$</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original algorithm</td>
<td>723</td>
<td>18</td>
<td>42</td>
<td>23</td>
<td>640</td>
<td>.44(.29-.60)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Modified algorithm (i)</td>
<td>723</td>
<td>25</td>
<td>74</td>
<td>16</td>
<td>608</td>
<td>.61(.45-.75)</td>
<td>.016</td>
<td></td>
</tr>
<tr>
<td>Modified algorithm (ii)</td>
<td>723</td>
<td>27</td>
<td>86</td>
<td>14</td>
<td>596</td>
<td>.66(.49-.79)</td>
<td>.008</td>
<td></td>
</tr>
<tr>
<td>Screening question</td>
<td>723</td>
<td>29</td>
<td>114</td>
<td>12</td>
<td>568</td>
<td>.71(.54-.83)</td>
<td>.003</td>
<td></td>
</tr>
</tbody>
</table>

*a: true positives, b: false positives, c: false negatives, d: true negatives, ppv: positive predictive value, npv: negative predictive value, LR: likelihood ratio. *p-values are corrected for multiplicity using Bonferroni-Holm procedure. a All of the first four questions are answered with “yes,” and presence of four or more somatic symptoms during an anxiety attack. b At least three of first four questions are answered with “yes,” other coding criteria unchanged. c At least two of first four questions are answered with “yes,” other coding criteria unchanged. d “In the last four weeks, have you had an anxiety attack—suddenly feeling fear or panic?”

### Table 3. Test characteristics of PHQ-PD of patients with a positive score on the depression or somatisation module of the PHQ.

<table>
<thead>
<tr>
<th>Cut off point</th>
<th>Total</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>sens</th>
<th>spec</th>
<th>ppv</th>
<th>npv</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive score on PHQ-depression module:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original algorithm</td>
<td>91</td>
<td>10</td>
<td>12</td>
<td>7</td>
<td>62</td>
<td>.59(.33-</td>
<td>.84(.73-</td>
<td>.45(.25-</td>
<td>.90(.80-</td>
</tr>
<tr>
<td>algorithm a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.81)</td>
<td>.91)</td>
<td>.67)</td>
<td>.95)</td>
</tr>
<tr>
<td>Screening question</td>
<td>91</td>
<td>13</td>
<td>38</td>
<td>4</td>
<td>36</td>
<td>.76(.50-</td>
<td>.49(.37-</td>
<td>.25(.15-</td>
<td>.90(.75-</td>
</tr>
<tr>
<td>question b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.92)</td>
<td>.60)</td>
<td>.40)</td>
<td>.97)</td>
</tr>
<tr>
<td><strong>Positive score on PHQ-somatisation module:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original algorithm</td>
<td>150</td>
<td>14</td>
<td>17</td>
<td>8</td>
<td>111</td>
<td>.64(.41-</td>
<td>.87(.79-</td>
<td>.45(.28-</td>
<td>.93(.87-</td>
</tr>
<tr>
<td>algorithm a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.82)</td>
<td>.92)</td>
<td>.64)</td>
<td>.97)</td>
</tr>
<tr>
<td>Screening question</td>
<td>150</td>
<td>17</td>
<td>44</td>
<td>5</td>
<td>84</td>
<td>.77(.54-</td>
<td>.66(.57-</td>
<td>.28(.18-</td>
<td>.94(.87-</td>
</tr>
<tr>
<td>question b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.91)</td>
<td>.74)</td>
<td>.41)</td>
<td>.98)</td>
</tr>
</tbody>
</table>

*a: true positives, b: false positives, c: false negatives, d: true negatives, ppv: positive predictive value, npv: negative predictive value, LR: likelihood ratio. a All of first four questions are answered with “yes,” and presence of four or more somatic symptoms during an anxiety attack. b “In the last four weeks, have you had an anxiety attack—suddenly feeling fear or panic?”*
Prevalence of panic disorder and the psychometric properties of the PHQ-PD

The prevalence of panic disorder (according to the SCID-I) in the total high-risk population (calculated with inverted probability weighing) was 5.7% (41 of 723 patients) with a 95% confidence interval of 4.1–7.7% (see Table 1).

The most widely used cutoff point to indicate a positive case for panic disorder is the algorithmic cutoff point. Of the 41 SCID-I-panic disorder patients 18 patients had a positive score on the algorithm of the PHQ-PD. Of the 682 SCID-I non-panic disorder patients 640 patients scored negative on this algorithmic cutoff point. This resulted in a PHQ-PD sensitivity of 0.44 (95% CI: 0.29–0.60), a specificity of 0.94 (95% CI: 0.92–0.95), a positive predictive value of 0.30 (95% CI: 0.19–0.43), and a negative predictive value of 0.97 (95% CI: 0.95–0.98).

The two modified algorithms (i and ii) generated significantly higher sensitivities compared to the original algorithm (.61(.45–.75); p-value 0.016 versus .66(.49–.79); p-value 0.008) and significantly lower specificities (.89(.87–.91); p-value <0.001 versus .87(.85–.90); p-value <0.001).

According to our analysis using only the screening question is significantly more sensitive compared to the original algorithm (p-value 0.003) but significantly less specific (p-value ≤0.001). This question had a sensitivity of 0.71 (95% CI: 0.54–0.83), a specificity of 0.83 (95% CI: 0.80–0.86), a positive predictive value of 0.20 (95% CI: 0.14–0.28), and a negative predictive value of 0.98 (95% CI: 0.96–0.99).

For details about the modified algorithms see Table 2.
Patients that scored positive on the PHQ-depression module or PHQ-somatisation module

Of the 723 patients 91 scored positive on the PHQ-depression module. The adjusted prevalence of panic disorder in this subgroup of 91 patients was 18.6% (95% CI: 11.6–28.5). Sensitivity and specificity figures for the diagnostic algorithm were 0.59 (95% CI: 0.33–0.81) and 0.84 (95% CI: 0.73–0.91) respectively with a positive predictive value of 0.45, and a negative predictive value of 0.90. Sensitivity and specificity figures for the single screening question were 0.76 (95% CI: 0.50–0.92) and 0.49 (95% CI: 0.37–0.60) respectively with a positive predictive value of 0.28 (95% CI: 0.18–0.41), and a negative predictive value of 0.94 (95% CI: 0.87–0.98) (Table 3).

Of the 723 patients 150 scored positive on the PHQ-somatoform disorder module. The adjusted prevalence of panic disorder in this subgroup of 150 patients was 14.7%. Sensitivity and specificity figures for the diagnostic algorithm were 0.64 (95% CI: 0.41–0.82) and 0.87 (95% CI: 0.79–0.92) respectively with a positive predictive value of 0.45, a negative predictive value of 0.93, a positive likelihood ratio of 4.8 (95% CI: 2.8–8.3), and a negative likelihood ratio of 0.4 (95% CI: 0.2–0.7). Sensitivity and specificity figures for the screening question were 0.77 (95% CI: 0.54–0.91) and 0.66 (95% CI: 0.57–0.74) respectively with a positive predictive value of 0.28, a negative predictive value likelihood ratio of 2.2 (95% CI: 1.6–3.1), and a negative likelihood ratio of 0.3 (95% CI: 0.2–0.8) (Table 3).

Discussion

The aim of our study was to investigate the test characteristics of PHQ-panic disorder (PHQ-PD) in a high-risk population in primary care. Our secondary aim was to test whether modifications in the questionnaires’ evaluation algorithms led to better operating characteristics. Finally we aimed to study the influence of a positive score on other modules of the PHQ (PHQ-depression and PHQ-somatisation) on the test characteristics of the PHQ.

In this high-risk population we found a prevalence of panic disorder of 5.7%. Regarding the literature this prevalence is higher than the prevalence in the unselected primary care population. The prevalence of panic disorder in the general population is about 1.5% and in the unselected primary care population about 4%. Higher rates of 20–30% have been described in subpopulations with gastro-intestinal or cardial complaints but these figures represented mostly lifetime prevalence and not point prevalence as we did. Furthermore none of these
prevalence studies used the strict criteria of the SCID-I to measure the prevalence of PD.

We conclude that the PHQ-PD can be recommended as a valid screening instrument. As was stated before the modified algorithms and the screening question of the PHQ-PD lead to an additional improvement in screening properties compared to the original algorithm.\textsuperscript{17} In contrast to Lowe et al we found a significant improvement of sensitivities of both modified algorithms.\textsuperscript{17} However the screening question showed the best sensitivity (screening question: sensitivity 0.71 (95\%CI 0.54–0.83) versus original algorithm: 0.44 (95\%CI 0.29–0.60); p-value 0.003). For screening purposes requiring high sensitivity, we endorse the screening question. For epidemiological purposes where disease prevalence is calculated we advise to use the original algorithm with the highest positive predictive value (0.30), combined with an additional diagnostic (DSM-IV based) clinical interview. Otherwise 70\% of patients that scored positive on the PHQ-PD will be wrongly labelled as having panic disorder.

The psychometric properties of the PHQ-PD in our study (n = 731) were similar to Becker et al which was performed in the unselected primary care population (n = 173), using the same reference standard as we did (SCID-I) (sensitivity .47 and specificity .96).\textsuperscript{36} On the other hand our results differed from other studies. Mazzotti et al reported a sensitivity of .17 and a specificity of .97 in dermatological patients (n = 170), also using the SCID-I as a reference standard.\textsuperscript{18} The authors mentioned translation problems of the Italian version of the PHQ as an explanation for the low sensitivity. Lowe et al. reported a sensitivity of .75 and a specificity of .96 in outpatients (n = 499), including medical, psychosomatic and primary care patients, using the SCID-I as a reference standard.\textsuperscript{17} We found lower sensitivities than Lowe et al. in our study population that consisted of a high-risk population in primary care. In other words, patients with panic disorder according to the SCID-I in our population showed less frequently a positive score on the PHQ-PD compared to patients from this other study.\textsuperscript{17} It might be possible that some patients with panic disorder from our high-risk population (consisting of many patients with physical and/or somatoform complaints) had more often a somatic or physical explanation for their panic attacks. These patients might not have recognised the PHQ-description of panic attack and scored therefore negative on the PHQ-PD. A general explanation for the variability of sensitivities among all studies is the relatively low prevalence of panic disorder which causes wide confidence intervals around sensitivity figures. The specificity in our study (.94 (95\%CI: .92–.95) was similar to other studies.

As could be expected, many patients scored also positive on the PHQ-depression module and/or somatisation module; of 41 patients with panic disorder 17 (41\%)
patients scored positive on the PHQ-depression module and 22 (54%) scored positive on the PHQ-somatization module. Scoring positive on the PHQ-depression or somatisation module increased the sensitivity of the PHQ-PD, but decreased its specificity. This is probably due to increased severity (and increased amount of symptoms) of panic disorder when patients are also suffering from depressive or somatoform symptoms. In addition, this could also be caused by overlapping symptoms (for instance ‘shortness of breath’ or ‘feeling your heart pounce’ could be symptoms of somatisation as well as panic attack). The positive predictive value increased because of the higher prevalence of panic disorder in these subgroups. For diagnostic purposes, the PHQ-PD performs better in subgroups with positive scores on the depression and/or somatisation module, making it a somewhat more useful instrument in clinical practice, where many patients are suffering from a combination of symptoms.

**Methodological strengths and limitations**

Our study was conducted in a large group of primary care patients using a standardized diagnostic instrument (SCID-I) that had been validated extensively as reference standard. The two-stage design proved efficient, without influencing statistical preciseness. On the other hand, this design resulted into an inclusion of only 41 patients with panic disorder. Confidence intervals for sensitivity were therefore as wide as previous studies. Larger studies are needed in combination with pooling meta-analyses to come to more preciseness of the sensitivity of this instrument.

Younger patients and male patients were less inclined to participate, probably due to less interest in a screening program for psychiatric disorders and lack of time because of a busier daily life. Patients belonging to the MHP group probably had already discussed their psychosocial problems with the GP during regular consultations and therefore rejected participation. The fact that the distribution among high-risk groups of patients who participated and patients who gave informed consent but could not be reached by telephone did not differ makes a relevant selection bias at this stage unlikely.

**Implications**

Our study is the first study of the PHQ-PD within specific high-risk groups for psychiatric disorders in general practice and the first study to account for positive scores on other modules than the PHQ-PD. The screening question of the PHQ-PD proved to be appropriate for screening purposes, considering its sensitivity; 71% of all patients with panic disorder scored positive on the screening question. We conclude that the screening question of the PHQ-PD can be used to screen for panic disorders in a high-risk population, but screening should be followed by a formal diagnostic procedure. Because of the low positive predictive value, two
(algorithm) or three (one screening question) of every four screen positive patients will not suffer from PD. Furthermore, we conclude that scores on other modules influence the test characteristics of the PHQ-PD. Clinicians should also be aware that the sensitivity and specificity of a screening module is dependent of the scores on other modules of the PHQ.
References


