Screening for binge eating disorders using the Patient Health Questionnaire in a community sample.

Ruth Striegel Weissman

Available at: https://works.bepress.com/ruth_striegel/65/
Screening for Binge Eating Disorders Using the Patient Health Questionnaire in a Community Sample

Ruth H. Striegel-Moore, PhD1*
Nancy Perrin, PhD2
Lynn DeBar, PhD2
G. Terence Wilson, PhD3
Francine Rosselli, PhD1
Helena C. Kraemer, PhD4

ABSTRACT
Objective: To examine the operating characteristics of the Patient Health Questionnaire eating disorder module (PHQ-ED) for identifying bulimia nervosa/binge eating disorder (BN/BED) or recurrent binge eating (RBE) in a community sample and to compare true positive (TP) versus false positive (FP) cases on clinical validators.

Method: Two hundred and fifty-nine screen-positive individuals and a random sample of 89 screen negative cases completed a diagnostic interview. Sensitivity, specificity, and positive predictive value (PPV) were calculated. TP and FP cases were compared using t-tests and Chi-square tests.

Results: The PHQ-ED had high sensitivity (100%) and specificity (92%) for detecting BN/BED or RBE, but PPV was low (10 or 19%). TP and FP cases did not differ significantly on frequency of subjective bulimic episodes, objective overeating, restraint, on BMI, and on self-rated health.

Discussion: The PHQ-ED is recommended for use in large populations only in conjunction with follow-up questions to rule out cases without objective bulimic episodes. © 2009 by Wiley Periodicals, Inc.

Keywords: Patient Health Questionnaire; bulimia nervosa; binge eating disorder; recurrent binge eating; screening; obesity

Introduction
Bulimic spectrum disorders [i.e., threshold or sub-threshold bulimia nervosa (BN) and binge eating disorder (BED)] have been shown to be associated with psychiatric comorbidity as well as health and psychosocial impairment.1-5 Although effective treatments are available,6-9 these eating disorders often go undetected and untreated,10-12 which has prompted experts to call for early identification through outreach and screening efforts.12

A myriad of assessment tools has been developed to measure eating pathology, yet evidence of their utility for case detection is limited and no uniformly accepted screening tool for eating disorders has emerged.13 Many instruments emphasize symptoms characteristic of anorexia nervosa and most were developed before BED was introduced into the Diagnostic and Statistical Manual (DSM) as a disorder in need of further study. The Eating Disorder Examination interview (EDE)14 is widely regarded as the “gold standard” measure for making an eating disorder diagnosis,15 and its self-report questionnaire version, the EDE-Q16 has gained increasing popularity for measuring eating disorder symptoms in both clinical and epidemiological studies because of the significant time burden associated with the interview version. Mond17,18 recently reported on the utility of the EDE-Q as a screener in community and primary care samples. Although the reported values for specificity, sensitivity, and positive predictive value (PPV) were favorable, the instrument’s length of 35 items makes it still quite burdensome for use as a screener.

Eating disorder screens are particularly needed for primary care settings for several reasons. One, although they typically do not receive treatment...
specifically for the eating disorder, studies have shown that individuals with an eating disorder tend to use more health services than individuals without an eating disorder\textsuperscript{19} and for most patients primary care providers represent the “point of entry” into the health care system.\textsuperscript{20} Hence, primary care providers play a pivotal role in treating patients or referring patients into specialty treatment. Two, prior research has shown that patients prefer being treated for their eating disorder in the primary care setting.\textsuperscript{12} Three, evidence suggests that primary care providers often do not detect the eating disorder (or, for that matter, other mental health problems), possibly at least in part because of the time pressure under which the providers have to deliver their health care services.\textsuperscript{11}

An important consideration in choosing a diagnostic screening instrument for use in a medical setting is the ability to balance respondent and medical staff burdens with the accuracy of the instrument. Spitzer et al.\textsuperscript{21} advocated use of a self-report questionnaire (Patient Health Questionnaire; PHQ) they developed based on a general screening interview for common mental disorders (PRIME-MD). Composed of brief modules assessing the most common mental disorders, the PHQ was designed to permit selective use of modules if one is interested in diagnosing a particular disorder rather than screening for multiple disorders. Some of the modules (e.g., mood disorders) have been subjected to extensive testing\textsuperscript{22–24} and translated into numerous languages.\textsuperscript{25,26}

Notwithstanding the widespread use of the PHQ for general screening for mental disorders in general, or for mood disorders in particular, few studies have used the Patient Health Questionnaire eating disorder module (PHQ-ED) as a screening tool. The sensitivity (89%) and specificity (96%) of this screening module was determined based on a subset of patients (number not reported) using the PRIME-MD interview for validation.\textsuperscript{21} One aim of the present study was to determine the operating characteristics of the PHQ-ED module when used for screening a sample of members of a large health maintenance organization. The operating characteristics of the PHQ-ED module when used for screening a sample of members of a large health maintenance organization have been continuously insured by the health plan for the 12 months prior to selection.

Participants and Procedures

Potential participants were randomly selected from the electronic health record (EMR) system of a large health maintenance organization, using criteria specified below. Excluded from the recruitment were individuals whose medical records indicated severe cognitive impairment (e.g., mental retardation), or those currently receiving treatment for a severe physical illness such as cancer. Also excluded were \textasciitilde100 plan members whose records indicated that they had preemptively opted out of any involvement in research. Potential participants had to have been continuously insured by the health plan for the 12 months prior to selection.

Figure 1 shows the flow of participants through the recruitment process. Beginning in June 2004 and ending in July 2005, letters were sent to male and female health plan members between 18 and 35 years of age [the population of adults for whom community epidemiological studies suggest the highest rates of binge eating disorders\textsuperscript{1–2,27} inviting them to complete a brief self assessment about their eating habits, body shape, and weight concerns, and other health and demographic factors. Participants who met PHQ-ED criteria for BN, BED, or RBE were invited to participate in an in-depth assessment using a state-of-the-art diagnostic interview. To examine the psychometric properties of the PHQ-ED in this population, a random sample of individuals who screened negative was also recruited to complete the diagnostic assessment. The final sample consisted of a total of 348 participants, 259 screen positive cases, and 89 screen negative cases. These participants were primarily white (87%), non-Hispanic (93%), women (82%), with at least some college experience (80%), and an average age of 28.18 (SD = 5.38). Participants who completed the diagnostic assessment did not differ from those who did not

Method

Participants and Procedures

Potential participants were randomly selected from the electronic health record (EMR) system of a large health maintenance organization, using criteria specified below. Excluded from the recruitment were individuals whose medical records indicated severe cognitive impairment (e.g., mental retardation), or those currently receiving treatment for a severe physical illness such as cancer. Also excluded were \textasciitilde100 plan members whose records indicated that they had preemptively opted out of any involvement in research. Potential participants had to have been continuously insured by the health plan for the 12 months prior to selection.

Figure 1 shows the flow of participants through the recruitment process. Beginning in June 2004 and ending in July 2005, letters were sent to male and female health plan members between 18 and 35 years of age [the population of adults for whom community epidemiological studies suggest the highest rates of binge eating disorders\textsuperscript{1–2,27} inviting them to complete a brief self assessment about their eating habits, body shape, and weight concerns, and other health and demographic factors. Participants who met PHQ-ED criteria for BN, BED, or RBE were invited to participate in an in-depth assessment using a state-of-the-art diagnostic interview. To examine the psychometric properties of the PHQ-ED in this population, a random sample of individuals who screened negative was also recruited to complete the diagnostic assessment. The final sample consisted of a total of 348 participants, 259 screen positive cases, and 89 screen negative cases. These participants were primarily white (87%), non-Hispanic (93%), women (82%), with at least some college experience (80%), and an average age of 28.18 (SD = 5.38). Participants who completed the diagnostic assessment did not differ from those who did not
complete the assessment on BMI \((t(610) = 1.25, p = .21)\), race \((\chi^2 < 1)\), or Hispanic ethnicity \((\chi^2 < 1)\). Participants who completed the assessment were older and reported better health \((M = 28.18 \text{ years}, SD = 5.38; M = 29.94, SD = 1.10)\) than those who did not complete the assessment \((M = 27.28 \text{ years}, SD = 5.62; M = 3.24, SD = 1.02), t(626) = 2.04, p < .04, \text{Cohen's } d = 0.16, \text{ and } t(620) = 3.56, p < .001, \text{Cohen's } d = 0.28, \text{ respectively. Eighty percent of those who completed the diagnostic assessment reported at least some college attendance compared to 70% of those who did not complete the assessment } [\chi^2(1) = 8.91, p < .003, \text{NNT} = 9.71]. \]

**Measures**

**Initial Screening Questionnaire.** The PHQ-ED\(^{21}\) is composed of six binary (yes/no) response items concerning binge eating and compensatory behaviors asked of everyone and up to two follow-up questions asked when binge eating or purging are endorsed. To assess binge eating, participants were asked to indicate if they often feel that they cannot control what or how much they eat (loss of control) and if they often eat, within any 2-h period, what most people would regard as an unusually large amount of food (overeating). Participants responding “yes” to both of these items are then asked to indicate if the behaviors in question occurred as often as twice a week in the last 3 months. Focusing on the past 3 months, respondents are asked to indicate whether they vomited, took laxatives, fasted, or exercised in order to avoid gaining weight after binge eating and, if yes, whether any of these behaviors occurred at least as often as twice a week on average during the past 3 months.

The PHQ-ED was augmented with three questions to further describe the eating pathology and general health of our sample, as well as with items assessing height, weight, overall health, and demographic information (date of birth, gender, race, ethnicity, education, and marital status). Experts have called for a re-evaluation of the frequency criterion for binge eating for making a diagnosis of BN or BED.\(^5\) To identify participants who engage in binge eating behaviors at a frequency below the diagnostic threshold, the PHQ-ED was expanded to include an item asking participants to indicate if the binge eating occurred as often as once a week in the past 3 months. An additional item, adapted from the EDE-Q,\(^{16}\) was included to assess dietary restriction. Participants indicated (yes/no) whether they consciously tried to limit the amount of food eaten in order to influence their shape or weight over the prior 3 months. Finally, participants were asked to rate their general health (1 = excellent, 5 = poor) over the prior 3 months. Binary demographic categories were created as follows: White (yes/no), Hispanic (yes/no), currently married or partnered (yes/no), and high-school graduate or less versus some college or more.

**Diagnostic Assessment.** The EDE,\(^{14}\) 12th edition with text edits from the 14th and 15th editions, was administered by telephone to validate eating disorder diagnoses. This standardized, semistructured, investigator-based interview measures the presence and severity of the core clinical features of eating disorders and is considered the “gold standard” method for assessing ED psychopathol-
ogy. Specifically, the diagnostic items of the interview were used to determine whether participants met DSM-IV criteria for BN, BED, or study criteria for RBE. Interviewers were blind to participant responses on the PHQ-ED and received extensive training, including 20 h of workshops with an expert EDE trainer as well as interviews with pilot participants. Interviewers continued practice interviews until they achieved three consecutive interviews with 100% expert agreement. Ongoing biweekly supervision continued throughout the data collection period. Five percent of the EDE interviews were randomly selected to be coded by one of the two most experienced assessors to determine reliability. Interrater reliability was very high for EDE diagnoses (κ = 0.961) as well as assessment of objective bulimic episode (OBE) days (ICC = 0.997) and episodes (ICC = 0.999) during the past 28 days.

**Analysis**

The software package SPSS (15.0) was used for all analyses. The sensitivity, specificity, and predictive values of the PHQ-ED were calculated based on the original PHQ-ED items. Sensitivity ("TP rate") refers to the probability that an individual would screen positive if a BED is actually present (as determined by EDE interview). Specificity ("true negative rate") is the probability that an individual would screen negative if a BED is not present. The PPV represents the probability that an individual classified as a case by the screener is also classified as a case by the EDE interview, and negative predictive value (PVN) that an individual classified as a noncase by the screener remains a noncase in the interview. Sensitivity and specificity were calculated adjusting for the sampling; as discussed by Kraemer, it is important to take the sampling into account, particularly for low-prevalence conditions like eating disorders. Specifically, because not all those that were screened went on to complete the EDE, sensitivity and specificity need to be adjusted to reflect the values for the population. The proportion that screened positive in the population (Q, the number who screened positive on the PHQ divided by the number screened) can be estimated directly from the data as can the PPV and PVN.

However, the prevalence (P), number of TPs (TP), true negatives (TN), FP, and false negatives (FN) must be adjusted using the following formulas:

\[ P = (Q \times PPV) + (Q' \times PVN') \]

\[ TP = Q \times PPV \]

\[ TN = Q' \times PVN \]

\[ FP = Q \times PPV' \]

\[ FN = Q' \times PVN' \]

Sensitivity is then TP/P and Specificity is TN/(1 − P).

In an effort to describe individuals who self-reported binge eating on the PHQ-ED but failed to qualify for a BED diagnosis, we compared TP participants and FP participants on measures of the frequency of three types of eating episodes, as measured by the EDE: OBEs (eating episodes involving both a sense of loss of control and eating objectively large amounts of food), subjective bulimic episodes (SBEs, eating episodes involving a sense of loss of control while eating amounts of food that are experienced by the participant as "large" but deemed by research criteria as below the threshold required for OBEs), and objective overeating (OO, eating episodes involving consumption of an objectively large amount of food in the absence of a sense of loss of control). The two groups also were compared on eating restraint, BMI, and self-rated health. Cohen’s d was used to estimate effect size.

### Results

Table 1 shows the distribution of participants across the possible screen (PHQ-ED) and interview (EDE) outcomes. The operating characteristics of the PHQ-ED as a screen for identifying BN/BED or RBE (which includes BN/BED) were as follows. For either diagnostic outcome, the sensitivity was 100%; specificity was 91.7% (BN/BED) and 92.4% (RBE). Although our analyses suggest that the probability that a person does not have an eating disorder given a negative screen is 100%, the probability that a person meets interview assessed criteria for BN/BED or RBE given a positive screen is fairly low (PPV = 0.104 and PVV = 0.185).

As shown in Table 2, FP cases had significantly fewer days with OBEs (a strong effect) than TP cases. The two groups did not differ significantly on number of days with SBEs, episodes of objective overeating, mean BMI, or average self-rated health. Prevalence of dietary restraint did not differ significantly [\( \chi^2(1) = 0.046, p = .83 \)]. A majority of FP cases (83.0%) and TP cases (81.6%) indicated that

**Table 1.** Distribution of participants across PHQ-ED and EDE outcomes

<table>
<thead>
<tr>
<th></th>
<th>EDE Outcome</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BED or BN</td>
<td>RBE</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>PHQ-ED Screen +</td>
<td>27</td>
<td>232</td>
</tr>
<tr>
<td>PHQ-ED Screen −</td>
<td>0</td>
<td>89</td>
</tr>
</tbody>
</table>

| PHQ-ED screen +, participant endorsed overeating with loss of control at least once per week over the last month; PHQ-ED screen −, participant reported neither overeating and/or loss of control or endorsed these behaviors but not at the required minimum threshold of once per week over the last month.
in the past 3 months they had been consciously trying to limit the amount of food they ate to influence their shape or weight.

Discussion

Epidemiological studies have shown that eating disorders involving binge eating as the core behavioral symptom (i.e., bulimia nervosa, BED, and their subthreshold variants) are the most common forms of eating disorders\(^1,\)\(^2,\)\(^27\) and that these eating disturbances are associated with personal and social suffering\(^5\). To our knowledge, this is the first systematic study of the psychometric properties of the eating disorder module of the PHQ for detecting such disorders. The PHQ-ED has the advantage of being brief (thus taking up little of the participants’ time) and readily available at no cost. We recruited a random sample of young adults, the age group most likely to experience a binge eating disorder,\(^2,\)\(^27\) from a large health maintenance organization with a socioeconomically diverse membership. Our findings suggest that the PHQ-ED is highly sensitive: not a single person among those who screened negative was found to experience RBE or meet full-syndrome criteria for BN or BED or RBE on the Eating Disorder Examination (EDE); false positive cases screened positive for a binge eating disorder on the PHQ-ED but were found to not meet diagnostic criteria for a binge eating disorder on the EDE.

Notwithstanding the group differences in the frequency of objective bulimic episodes, the two

| TABLE 2. Comparison of true positive cases and false positive cases on number of days with binge eating or overeating episodes, body mass index (BMI), and self-rated health |
|-----------------|-----------------|-----------------|-----------------|
| **True Positive Mean (SD)** | **False Positive Mean (SD)** | **t(df)** | **p-value** | **Cohen’s d** |
| OBE | 11.67 (7.88) | 0.42 (1.46) | t(257) = 19.43 | <.0001 | 1.99 |
| SBE | 5.70 (7.54) | 4.02 (6.62) | t(255) = 1.52 | .13 | 0.24 |
| OO | 0.95 (3.34) | 0.76 (2.54) | t(256) = 0.46 | .65 | 0.07 |
| BMI | 31.76 (8.84) | 31.92 (7.85) | t(253) = -0.13 | .90 | -0.02 |
| Health | 3.35 (1.08) | 3.10 (1.05) | t(254) = 1.50 | .14 | 0.23 |

Note: OBE, objective bulimic episode; SBE, subjective bulimic episode; OO, objective overeating; true positive cases screened positive for a binge eating disorder [i.e., bulimia nervosa (BN), binge eating disorder (BED), or recurrent binge eating (RBE)] on the PHQ-ED and were found to meet diagnostic criteria for BN or BED or for RBE on the Eating Disorder Examination (EDE); false positive cases screened positive for a binge eating disorder on the PHQ-ED but were found to not meet diagnostic criteria for a binge eating disorder on the EDE.
groups did not differ on BMI or self-rated health, two commonly used indicators of clinical significance. Furthermore, in both groups, most participants reported high levels of dietary restraint. The average BMI values suggest that these participants were overweight or obese. Thus, the PHQ-ED appears to identify individuals who might benefit from interventions focused on weight management.

Several limitations of our study need to be acknowledged. Recruitment rates were disappointingly low in this epidemiological phase when participants were invited from the membership rosters rather than being approached more directly through primary care clinics and we were unable to address the question of systematic recruitment biases at the initial recruitment phase. Among participants invited for confirmatory interview assessment, those agreeing to be interviewed were older, better educated, and reported to be in better health. Although statistically significant, effect size estimates for these differences were small, however. Our sample included predominantly white individuals; our findings therefore cannot be generalized to ethnic minority samples. These limitations notwithstanding, our study offers the first systematic effort to test whether the PHQ, a widely used screener, may be useful for identifying eating disorders in the community. Our answer to this question is a qualified “yes,” because despite its favorable specificity, the PHQ-ED likely will yield a very high FP rate when used among non-treatment-seeking populations. Additional research is needed to determine whether the PPV of this or other screening tools can be enhanced, for example, by offering specific definitions of what constitutes binge eating.32,33

The contents of this study are solely the responsibility of the authors and do not necessarily represent the official of the NIH, NIMH, NIDDK, or the Kaiser Foundation Research Institute.

References