Convergent Validity of a Single Question with Multiple Classification Options for Depression Screening in Medical Settings

H. Edward Fouty, Hanny C. Sanchez, Daniel S. Weitzner, Brianna M. Brandon, Rachel A. Mills, Estefany S. Bologna, Daniel Guzman, and Nicole A. Baker

Abstract— **The purpose of this study was to assess the convergent validity of a single depression question with multiple classification options for depression screening. Participants were 40 medical inpatients. The age range of our sample was 18 to 85 years (M = 56.15, SD = 17.66). A clinical interview and the BDI-II were administered. The correlation between patients' self-rating classification of depression and their BDI-II classification was significant, $r_s(38) = .90$, $p < .01$. Follow-up repeated-measures chi-square revealed a statistically significant association between BDI-II classification and patients' self-rating classification, $\chi^2(9, N = 40) = 47.79$, $p < .005$. Significant positive standardized residuals revealed a clear linear relationship between BDI-II and patient self-rating classifications. Our data support the use of a single depression question with multiple classification options as a useful and valid means of quickly screening for the presence of depression by frontline health care professionals.**

Keywords-convergent validity; depression screening; single question; BDI-II; severity level

I. INTRODUCTION

Depression is one of the most common mental illnesses globally in both medical and non-medical populations. The World Health Organization noted that in 17 countries 1 in 20 people reported having an episode of depression in the past year [1]. In the United States, the Centers for Disease Control and Prevention reported a 9.1% prevalence rate of current depression in the general population [2]. Reference [3] observed a lifetime prevalence of 16.6% for major depression with a lifetime morbid risk of 29.9% for persons 13 years of age and older. The prevalence of depression in medically ill persons has been reported to be significantly higher than healthy persons, with rates ranging from 20% to 40% [4]. As staggering as these rates may seem, in an examination of the existing literature, categorical prevalence rates of depression in patients with comorbid medical illnesses as high as 75% have been reported [5]. The importance of considering prevalence rates in medically ill persons lies in the finding that 47.6% of psychologists in the United States work in medical settings [6, 7]. It is therefore essential that frontline healthcare professionals working in medical settings be able to quickly and effectively evaluate and screen for depressive symptomology.

Depression is assessed through self-report questionnaires or a structured clinical interview (e.g., SCID-CV) [8]. With regard to psychometric options, the BDI-II [9] is one of the most commonly used [10, 11]. The BDI-II measures a patient's severity level of depression: none/minimal, mild, moderate, or severe. An evaluation of the psychometric properties of the BDI-II has demonstrated that this instrument yields reliable, internally consistent, and valid assessments of depression in medical care settings [12].

Due to the time-limited nature of assessment often observed in medical settings, efforts should be made to develop screening procedures that are valid, yet brief enough to be administered to medically ill persons. Previous studies have investigated the effectiveness of a one or two question format as a means of screening for depression in medical settings. A single depression question format entails comparing patient responses to a depression question (e.g., "Are you depressed?") to a clinical interview. In a two question format, the depression question is combined with a loss of interest question (e.g., "Have you experienced loss of interest in things or activities that you would normally enjoy?"). Patient responses are again typically compared to a diagnosis from a structured clinical interview.

Results are presented in terms of sensitivity and specificity. Sensitivity refers to the true positive rate; the degree of agreement between patients who describe themselves as depressed and a finding of depression on the secondary measure (typically a clinical interview). Specificity is the true negative rate; the degree of agreement between those who identify themselves as not depressed and a negative finding on the secondary measure [13].

Reference [14] conducted a meta-analysis of 22 studies from primary care utilizing a single question test. Results indicated that a single question of depression compared to a clinical interview yielded a sensitivity of 32%. Thus, 68% of the depressed patients were not correctly identified. Specificity was 97%. The authors concluded that a single item question was unacceptable if relied upon alone. A follow-up Bayesian meta-analysis [15], investigated the utility of using a one or two question format for detecting depression in medically ill (cancer) patients when compared to a clinical interview. A single depression question yielded a sensitivity of 72% and a

DOI: 10.5176/0000-0002_1.1.1

specificity of 83%. A single loss of interest question yielded a sensitivity of 83% and a specificity of 86%. A combined (two question) format (i.e., depression and loss of interest) revealed a sensitivity of 91% and a specificity of 86%. The author concluded that the two question format was significantly more accurate compared to a single question format, but neither format should be considered a definitive assessment of depression.

Several studies have been published since the meta-analyses described above. Reference [16] compared several depression instruments and the single question "Do you think you suffer from depression?" to a diagnosis made by a structured interview. An 83.3% sensitivity and an 82.9% specificity were observed. The authors noted that a single depression question is limited because it cannot assess depression severity. Reference [17] compared a single depression question to the Hopkins Symptom Checklist-25 (HSCL-25) [18]. Using an impressive receiver operating characteristic (ROC) curve analysis, the authors observed a sensitivity of 79.4% and a specificity of 80.8% for distinguishing not depressed versus depressed patient self-ratings. Reference [11] compared the mood section of the PRIME-MD [19] to a single depression question and the BDI-II (with cutoff scores of 14 and 16). The single question "Are you feeling depressed?" compared to the PRIME-MD revealed a sensitivity of 50% and a specificity of 94%. Comparison of a single question to the BDI-II (using the manualized cutoff score of 14) revealed a sensitivity and specificity of 90% and 64%, respectively.

Although different findings and conclusions have populated the scientific literature, the general consensus is that a single question for depression screening is unacceptable, but a two question screen has merit. However, to date there has been no investigation of the convergent validity of a single screening question with multiple classification options. The use of a multiple classification format addresses the lack of ability of single and two question formats to address depression severity. The purpose of the present study was to assess the convergent validity of a single question with multiple classification options to screen for depression severity of patients in medical settings.

II. Method

A. Participants

Our sample consisted of 40 medical inpatients at a large medical center in Central Florida. Admitting medical diagnoses varied. The age range of our sample was 18 to 85 years (M = 56.15, SD = 17.66). The sample consisted of 22 males between the ages of 22 and 85 years (M = 54.09, SD = 17.39) and 18 females between the ages of 18 and 79 years (M = 58.67, SD = 18.17). The education level (years completed) of our sample ranged from 8 to 20 years (M = 13.23, SD = 2.57); males 8 to 20 years (M = 14, SD = 3.01) and females 8 to 14 years (M = 12.28, SD = 1.49). Racial/ethnic demographics were: White (n = 39, 97.5%) and Black or African-American (n = 1, 2.5%).

B. Materials and Procedures

Following acquisition of informed consent, a clinical interview, consisting of background information and psychological diagnostic information based on the DSM-IV-TR criteria, was conducted. During the subjective complaints portion of the interview, patients were asked to rate their level of depression over the past two weeks ("How would you rate your level of depression: none/minimal, mild, moderate, or severe?"). The BDI-II was then self-administered. The modal time between patient self-rating of level of depression and administration of the BDI-II was between 10 and 15 minutes.

III. Results

Preliminary analysis of our data was undertaken to compare our findings to those of previous research on the use of a single question of depression. We examined the relationship between the patients' self-rating of depression classifications (depressed, not depressed) and their BDI-II classifications (depressed - scores 14-63, not depressed - scores 0-13). Of the 25 patients identifying themselves as depressed, 24 scored in the depressed range on the BDI-II, resulting in a sensitivity of 96%. Of the 15 patients who identified themselves as not depressed, 14 scored in the none/minimal range on the BDI-II, resulting in a specificity of 93.33%.

Of primary interest was the convergent validity of a single depression question with multiple classification options (i.e., none/minimal, mild, moderate, or severe). A Spearman rank-order correlation (rho) was conducted to assess the relationship between specific classifications of patients' self-rating of depression and their BDI-II classifications. A two-tailed test of significance revealed a significant positive relationship between self-rating level of depression and BDI-II classification, $r_s(38) = .90$, $p < .01$. A follow-up repeated-measures chi-square revealed a statistically significant association between BDI-II classification and patients' self-rating classification, $\chi^2(9, N = 40) = 47.79$, $p < .005$. Patients who self-rated their level of depression in a specific classification were significantly more likely to score in the same classification on the BDI-II. The observed frequencies and standardized residuals are presented in Table 1. An examination of the significant positive standardized residuals clearly reveals the pattern of responses. That is, there is a clear linear relationship between BDI-II and patient classifications.

IV. Discussion

While previous research has focused on sensitivity and specificity, neither of these values are appropriate for determining validity – a critical necessity in all diagnostic endeavors. The sensitivity and specificity of a single question format noted in prior studies have varied greatly [14]. A two question format (i.e., a depression question and a loss of interest question) has been reported to be more accurate, but still with limitations (e.g., inability to identify depression severity) [16]. The purpose of this study was to assess the convergent validity of a single depression question with

multiple classification options (i.e., "How would you rate your level of depression: none/minimal, mild, moderate, or

TABLE I. CELL OBSERVED FREQUENCIES (STANDARDIZED RESIDUALS) OF BDI-II X SELF-RATING SEVERITY CLASSIFICATIONS

	Self-rating Classification			
BDI-II Classification	***None/ Minimal***	***Mild***	***Moderate***	***Severe***
None/ Minimal	14 (3.5)	1 (-0.6)	0 (-2.1)	0 (-1.7)
Mild	1 (-0.4)	2 (2.1)	1 (-0.2)	0 (-0.9)
Moderate	0 (-2.0)	2 (0.5)	7 (2.0)	2 (-0.1)
Severe	0 (-1.9)	0 (-1.1)	4 (0.6)	6 (2.8)

severe?"). This approach is capable of identifying depression in conjunction with severity level.

The present data support the use of a single depression question with multiple classification options as a useful, accurate, and valid means of quickly screening for depression in medical settings. The significant association (r_s = .90) between self-rating of depression classification and depression classification on the BDI-II indicates that persons who rate themselves in a specific classification of depression are more likely to be rated in the same depressed classification on the BDI-II.

In terms of future research, we recommend authors correct for the minor limitations of our study by using a larger and more demographically diverse sample, including participants outside the United States. It may also be beneficial to include participants from various medical settings (i.e., primary care settings, clinics, etc.).

Many frontline health care professionals (i.e., hospital and community physicians, social workers, case managers, etc.) are often faced with the need to quickly screen for possible depression as it may affect the patient's medical care, health and discharge planning, and/or treatment interventions. However, these frontline health care professionals are not likely to have been rigorously trained in the administration and interpretation of psychometric instruments. Therefore, the use of a single question with multiple classification options allows such professionals to quickly and accurately screen for depression. We are not advocating that our approach for depression screening replace psychometric evaluations and clinical interviews by qualified mental health professionals (e.g., psychologist, psychiatrist). Our findings are presented as a valid and psychometrically sound screening to be used by frontline health care personnel. Should a positive finding of depression be observed, we recommend that patients be referred for a thorough psychological/psychiatric evaluation.

REFERENCES

[1] M. Marcus et al. (2012). *A global public health concern* [Online Article]. Available: http://www.who.int/mental_health/management/depression/who_paper_depression_wfmh_2012.pdf

[2] Centers for Disease Control and Prevention. (2011, Mar. 31). *An estimated 1 in 10 U.S. adults report depression* [Online Article]. Available: http://www.cdc.gov/features/dsdepression/

[3] R. C. Kessler et al., "Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States," *Int. J. of Methods in Psychiatric Research*, vol. 21, no. 3, pp. 169-184, Sep. 2012.

[4] R. G. Robinson and K. R. R. Krishnan, "Depression and the medically ill," in *Neuropsychopharmacology: The fifth generation of progress*, K. Davis, D. Charney, J. Coyle, & C. Nemeroff, Eds., 1st ed. Philadelphia: Lippincott Williams & Wilkins, 2002, ch. 81, sec. 8, pp. 1179-1185.

[5] D. L. Evans et al., "Mood disorders in the medically ill: Scientific review and recommendations," *Biological Psychiatry*, vol. 58, no. 3, pp. 175-189, Aug. 2005.

[6] American Psychological Association. (2011). *Careers in psychology: What psychologists do and where they do it* [Online Article] Available: http://www.apa.org/careers/resources/guides/careers.aspx?item=4

[7] D. Michalski et al. (2011, June). *2009 doctorate employment survey* [Online Article]. Available: http://www.apa.org/workforce/publications/09-doc-empl/

[8] *Structured Clinical Interview for DSM-IV Axis I Disorders,* American Psychiatric Press, Washington, DC, MD, 1996.

[9] Manual for the Beck Depression Inventory, 2nd ed., Psychological Corporation, San Antonio, TX, 1996.

[10] A. Aalto et al., "The Beck Depression Inventory and General Health Questionnaire as measures of depression in the general population: A validation study using the Composite International Diagnostic Interview as the gold standard," *Psychiatry Research*, vol. 197, no. 1-2, pp. 163-171, May 2012.

[11] F. Warmenhoven et al., "The Beck Depression (BDI-II) and a single screening question as screening tools for depressive disorder in Dutch advanced cancer patients," Supportive Care in Cancer, vol. 20, no. 2, pp. 319-324, Jan. 2011.

[12] R. C. Arnau, "Psychometric evaluation of the Beck Depression Inventory- II with primary care medical patients," *Health Psychology*, vol. 20, no. 2, pp. 112-119, Mar. 2001.

[13] R. Parikh et al. "Understanding and using sensitivity, specificity, and predictive values," *Indian J. of Ophthalmology*, vol. 56, no. 1, pp. 45-50, Jan.-Feb. 2008.

[14] A. J. Mitchell and J. C. Coyne, "Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies," *British J. of General Practice,* vol. 57, no. 535, pp. 144-151, Feb. 2007.

[15] A. J. Mitchell, "Are one or two simple questions sufficient to detect depression in cancer and palliative care? A Bayesian meta-analysis," British J. of Cancer, vol. 98, no. 12, pp. 1934-1943, May 2008.

[16] L. Ayalon et al., "Do you suffer from depression?" *Int. J. of Geriatric Psychiatry*, vol. 25, no. 5, pp. 497-502, Aug. 2009.

[17] S. E. Reme and H. R. Eriksen, "Is one question enough to screen for depression?" *Scandinavian J. of Public Health*, vol. 38, no. 6, pp. 618-624, Aug. 2010.

[18] L. R. Derogatis, "The Hopkins Symptom Checklist (HSCL-25): A self-report symptom inventory," *Behavioral Sci.*, vol. 19, no. 1, pp. 1-15, Jan. 1974.

[19] R. Fraguas et al., "The detection of depression in medical settings: A study with PRIME-MD," J. of Affective Disorder, vol. 91, no. 1, pp. 11-17, Nov. 2006.