The Communities That Care Brief Depression Scale: Psychometric Properties and Criterion Validity

Isaac C. Rhew[1,†], **Kathryn C. Monahan**[2], **Sabrina Oesterle**[3], and **J. David Hawkins**[3]

[1]Department of Psychiatry and Behavioral Sciences, University of Washington

[2]Department of Psychology, University of Pittsburgh

[3]Social Development Research Group, University of Washington

Abstract

For community-level approaches to preventing depression, measures of depression that are brief as well as valid are needed, particularly given competing demands in surveys for assessment of other outcomes including substance use, delinquency, and their associated risk factors. This study examined the validity of a four-item adolescent depression measure, the Communities That Care Brief Depression Scale (CTC-BDS). Data were from a survey of adolescents (N = 3,939) participating in the Community Youth Development Study (CYDS), a community-randomized trial involving 24 U.S. towns. The Patient Health Questionnaire nine-item (PHQ-9) was the criterion standard used to define major depressive disorder (MDD). Sensitivity and specificity of the CTC-BDS were both > .8 at a cutpoint of 6 and the area under the receiver operating characteristic curve was .91. Its strong psychometric properties and brevity make the CTC-BDS a useful measure for communities to monitor levels of youth depression.

Keywords

brief depression scale; community prevention; psychometrics; adolescent mental health

Background

Monitoring the occurrence of behavioral health problems among youth at the community level is a key element for effective public health practice. Community-level strategies can use epidemiologic data to inform the selection and implementation of suitable prevention programs. Although a number of community-wide strategies to improve youth outcomes have focused on reducing externalizing problems such as drug use and delinquent behaviors (Hawkins & Catalano, 2002; Spoth et al., 2007), very few attempt to monitor internalizing problems such as depression. To some extent, this is due to the fact that clinical mental health assessments, even for a single disorder, are time consuming, costly, and require substantial training for correct administration. Beyond these issues, however, community

[†]Corresponding author. University of Washington, Department of Psychiatry and Behavioral Sciences, Center for the Study of Health and Risk Behaviors, 1100 NE 45th St., #300, Seattle, WA 98195; rhew@u.washington.edu; telephone: +1 206 221-1897; FAX: 206 616-1705.

prevention planning does not seek to identify specific individuals in a population who are experiencing issues with mental health. In contrast, community prevention planning primarily seeks to determine the prevalence of depressive symptoms, or other symptomatology, within the community. As such, standard clinical assessment interviews are a poor match for the needs of community-based prevention planning efforts. Brief measures that can be administered community wide with limited cost would assist communities in evaluating the prevalence of depression in the community and help guide prevention programming. The needed measure must be accurate, validated for use within an adolescent population, and short enough for inclusion on school-based surveys, particularly given the competing desire to assess other outcomes such as drug use, delinquency, and associated risk factors through school-based surveys. This paper describes the results of an assessment of the validity of the four-item Communities That Care Brief Depression Scale (CTC-BDS) used in the Communities That Care (CTC) Youth Survey against a diagnosis of major depressive disorder (MDD) among a sample of adolescents.

Communities That Care (CTC) is a coalition-based prevention system that mobilizes community leaders and stakeholders to adopt a science-based approach to prevention that involves the local collection of epidemiologic data on risk and protective factors experienced by youth to inform selection and implementation of tested and effective programs to reduce the initiation and prevalence of problem behaviors (Hawkins & Catalano, 2002). Recent findings from a community-randomized trial have shown CTC to reduce risk for and prevalence of a number of problem behaviors among youth (Hawkins, Oesterle, Brown, Abbott, & Catalano, 2014; Hawkins et al., 2009; Hawkins et al., 2012). The CTC system uses the Communities That Care Youth Survey (CTCYS) to collect community data on the prevalence of risk and protective factors and problem behaviors among cross-sectional samples of youth (Arthur, Hawkins, Pollard, Catalano, & Baglioni, 2002). The survey includes a four-item CTC Brief Depression Scale (CTC-BDS) developed to assess depressive symptomology among youths in communities. Evidence suggests that the four-item CTC-BDS has strong construct validity, demonstrating expected correlations concurrently and longitudinally with risk and protective factors and problem behaviors (Monahan, Oesterle, Rhew, & Hawkins, 2014). Furthermore, in a sample of youth from Australia and Washington State, the CTC-BDS showed a high correlation (r = .77) with the Short Mood and Feelings Questionnaire (Angold et al., 1995), a commonly used epidemiologic depression scale for youth (Monahan et al., 2014). The present study is an assessment of the CTC-BDS against a diagnostic criterion standard to provide evidence regarding the criterion validity of this measure. Strong evidence of criterion validity would increase confidence in use of the scale for monitoring levels of depression among community youth to guide prevention efforts.

Methods

Data for this study came from the Community Youth Development Study (CYDS), a community-randomized trial of CTC in 24 incorporated towns across seven states (Hawkins et al., 2008). Communities were matched in pairs within state on population size, economic indicators, and crime rates. The towns participating in this study were small to moderate in population size with a range of approximately 1,500 to 50,000 residents. As part of this

randomized trial, a longitudinal panel of 4,407 fifth-grade students was followed annually, except for 11th grade, to age 19 (i.e., 1 year after the 12th-grade assessment). For this analysis, the age 19 survey was used because this was the only wave that included the depression criterion standard, the Patient Health Questionnaire 9-item (PHQ-9; described below), as well as the CTC-BDS. Students from both the intervention and control communities were included in the analysis because there was no evidence of differences in mean CTC-BDS or PHQ-9 scores by study condition. Among 4,399 still living individuals available for participation at age 19 who had not refused study participation at earlier waves, 3,986 (90.6%) participated in this survey. For this study, 47 additional participants were excluded because they were missing data on either the CTC-BDS or PHQ-9, yielding a final analysis sample of 3,939 adolescents.

Participants in the study completed a version of the CTCYS, the Youth Development Survey (YDS), which was slightly modified for use in the longitudinal panel. The majority of participants completed the survey online (89.2%), while others completed a paper-and-pencil questionnaire (10.8%). All study procedures were reviewed and approved by the University of Washington Institutional Review Board.

Measures

CTCYS Brief Depression Scale (CTC-BDS)—This measure asks participants to rate their agreement with four statements: "Sometimes I think life's not worth it."; "I think I am no good at all."; "I am inclined to think I am a failure."; and "In the past year, I have felt sad most days." Consistent with other items on the CTCYS, response options are on a 4-point ordinal scale: 0 ("NO"), 1 ("no"), 2 ("yes"), and 3 ("YES"). Items are summed to yield a total score where higher scores indicate a greater level of depressive symptomology.

Patient Health Questionnaire (PHQ-9)—The PHQ-9 was used as the criterion standard for MDD in this study. The PHQ-9 is a widely used measure to assess major depression in clinical and research settings (Kroenke, Spitzer, & Williams, 2001). It consists of nine items that map directly onto Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria for MDD. For each item, respondents are asked to respond how often the symptom bothered them during the past 2 weeks on a 4-point scale from 0 ("Not at all") to 3 ("Nearly every day"). A diagnosis of MDD is assigned based on endorsement of at least five of the nine symptoms occurring at least "more than half the days" or "nearly every day," with one of the endorsed symptoms being either "little interest or pleasure in doing things" or "feeling down, depressed or hopeless." A total score also can be calculated as the sum of the nine items for a possible range of 0 to 27. The PHQ-9 has shown strong criterion validity when assessed against clinician diagnostic interviews in the general population as well as other specific populations, including adolescents selected from primary care clinics (Richardson et al., 2010).

Analytic Plan

We first assessed the internal properties of the CTC-BDS. The internal consistency of the CTCYS-BDS was examined by calculating Cronbach's alpha in this sample. Because of the ordinal nature of the items, the estimation of the alpha was based on a polychoric correlation

matrix (Gadermann, Guhn, & Zumbo, 2012). To examine whether the four items of the scale represented a single dimension, we used Confirmatory Factor Analysis (CFA) to assess model fit for a one-factor solution. Fit statistics included the Root Mean Squared Error of Approximation (RMSEA), Tucker-Lewis Index (TLI) and the Comparative Fit Index (CFI). The items were specified as ordinal, and diagonally weighted least squares estimation was used.

The CTC-BDS was then evaluated against the depression criterion standard, the PHQ-9. Spearman's ρ was estimated for the correlation between the CTCYS-BDS and the PHQ-9 total scores. To assess the accuracy of the CTCYS-BDS against the MDD diagnosis, area under the receiver operating characteristic curve (AUC) and the corresponding 95% confidence interval were estimated. The receiver operating characteristic curve was generated by plotting sensitivity (the proportion of respondents classified as positive by the criterion standard that are classified as positive on the screening measure) against 1 – specificity (the proportion of respondents classified as negative on the criterion standard that also are classified as negative on the screening measure) across the possible CTC-BDS scores. AUC provides an estimate of the ability of the measure to correctly classify individuals as having the condition or not where an AUC of .5 indicates the test does not show any discrimination between a case and non-case beyond random chance, while a score of 1 indicates perfect discrimination. Recommended guidelines are that an AUC of less than .7 indicates low diagnostic accuracy, .7 to .9 reflects moderate accuracy, and greater than .9 indicates high accuracy (Henderson, 1993).

Sensitivity and specificity, as well as positive predictive value (PPV; the proportion of individuals at a given cutpoint score or higher who are also classified as positive on the criterion standard) and negative predictive value (NPV; the proportion of individuals who are classified as negative on the screening measure who are also classified as negative on the criterion standard) were calculated for each possible CTC-BDS score. As done elsewhere, we highlight the sensitivity and specificity at the nearest cutoff score where sensitivity and specificity intersect (Rhew et al., 2010). All analyses were conducted in R, version 3.0. Internal consistency was calculated using the "psych" package (Revelle, 2014), CFA was conducted using the "lavaan" package (Rosseel, 2012), and AUC analysis was conducting using the "pROC" package (Robin et al., 2011).

Results

In this sample of 3,939 individuals, the mean age was 19.2 years (SD = .44), 51.4% were female, 32.5% were non-White, and 20.4% were of Hispanic ethnicity. The mean CTC-BDS score was 3.09 (SD = 3.14). The mean total PHQ-9 score was 4.23 (SD = 5.08), and 257 (6.5%; 95% CI: 5.8–7.3) participants were classified as having a diagnosis of MDD which was slightly lower than the prevalence of 12-month major depression among 18- to 29-year-olds (8.3%) observed in the National Comorbidity Survey-Replication study (Kessler et al., 2003).

In this sample, internal consistency for the CTC-BDS was high (α = .93), which was consistent with estimates observed at earlier ages in this sample (Kroenke et al., 2001).

Further, CFA findings showed that factor loadings for each CTC-BDS item were high (standardized loadings > .8) and that the single-factor model showed strong model fit (RMSEA = .042; TLI = .999; CFI > .999), suggesting that the scale represented a unidimensional construct.

When comparing the CTC-BDS to the criterion standard, correlation between the CTC-BDS total score and the PHQ-9 total score was strong (Spearman's ρ = .62, p < .001). Further, as shown in the receiver operating characteristic curve (Figure 1), the CTC-BDS demonstrated high accuracy when assessed against MDD diagnosis (AUC = .91; 95% CI: .90, .93). The score where sensitivity and specificity were closest to intersecting was 6, and 22.4% of the sample met this threshold. At this cutpoint, sensitivity was .88 (95% CI: .83, .91), specificity was .82 (95% CI: .81, .83), PPV was .26 (95% CI: .24, .27), and NPV was .99 (95% CI: .99, .99). Data on sensitivity, specificity, PPV, and NPV at each CTC-BDS score are available upon request.

Discussion

Community-level prevention of depression among youth depends on a valid and brief assessment of depressive symptomology in the population to guide prevention program selection and ongoing monitoring of prevention progress. In this study of adolescents, the four-item CTC-BDS showed strong psychometric properties, suggesting its utility as a measure of depressive symptoms in community surveys where mental health measures must compete for time and space with assessments of other outcomes, including problem behaviors and risk factors. It showed strong internal consistency and evidence for unidimensionality. Importantly, the scale showed high accuracy when compared against a diagnosis of MDD as indicated by the high AUC estimate. Further, at a cutoff of 6, both sensitivity and specificity were high, indicating a good ability to distinguish the prevalence of cases of likely MDD and non-cases.

The fact that the scale demonstrated strong criterion validity is not surprising. The four items selected reflect symptoms found in the DSM. Further, three out of the four items in this measure reflect cognitive symptoms (life's not worth it, I'm no good, feel like a failure). Other studies using item response theory methods have found that these types of cognitive symptoms show the strongest discrimination compared to other depressive symptoms such as physical symptoms like decreased energy or changes in appetite (Sharp, Goodyer, & Croudace, 2006).

It is worth nothing that although sensitivity, specificity, and negative predictive value were all high at the cutpoint of 6, positive predictive value was fairly low. In this sample, 26% of those who scored a 6 or higher would have been classified as having MDD on the criterion standard which corresponds to a high false positive rate. However, given that the prevalence of current MDD in this population was relatively low (6.5%), extremely high values of specificity would be needed in order to obtain a high PPV. It should be noted that the CTC-BDS is not intended for individual screening or diagnostic purposes to identify individuals in need of mental health services. Instead, data from the CTC-BDS are intended to be used at

the community level to gauge the need for preventive mental health services in the population.

There were limitations to this study. First, the criterion standard was based on a nine-item measure rather than a diagnostic clinical interview. However, the PHQ-9 has shown very strong psychometric properties and criterion validity when assessing clinical diagnoses. The age range of this sample was restricted to adolescents approximately 19 years of age. Further, this study was conducted in a sample from small- and medium-sized towns across seven states. Thus, results from this study may not be generalizable to the general adolescent population.

As communities consider prevention strategies to address mental health problems such as depression among youth, it will be important to assess and continue to monitor levels of depression in the population using valid and brief measures that can be easily incorporated in community-wide surveys. Findings from this study support the use of the CTC-BDS as part of an epidemiologic assessment of depressive symptoms in an adolescent population. Using a cutpoint of 6 for the measure, communities could estimate the proportion of youth who are likely experiencing clinically significant depression and assess trends in prevalence over time. The results also suggest that the CTC-BDS is a valid tool for use in research studies examining the etiology, sequelae, and prevention of depressive symptoms.

Acknowledgments

References

Angold A, Costello EJ, Messer SC, Pickles A, Winder F, Silver D. Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. International Journal of Methods in Psychiatric Research. 1995; 5:237–249.

Arthur MW, Hawkins JD, Pollard JA, Catalano RF, Baglioni AJ Jr. Measuring risk and protective factors for substance use, delinquency, and other adolescent problem behaviors. The Communities That Care Youth Survey. Evaluation Review. 2002; 26:575–601. [PubMed: 12465571]

Gadermann AM, Guhn M, Zumbo BD. Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. Practial Assessment, Research and Evaluation. 2012; 17:1–13.

Hawkins, JD.; Catalano, RF. Investing in your community's youth: An introduction to the Communities That Care system. South Deerfield, MA: Channing Bete; 2002.

Hawkins JD, Catalano RF, Arthur MW, Egan E, Brown EC, Abbott RD, Murray DM. Testing Communities That Care: The rationale, design and behavioral baseline equivalence of the Community Youth Development Study. Prevention Science. 2008; 9:178–190. [PubMed: 18516681]

Hawkins JD, Oesterle S, Brown EC, Abbott RD, Catalano RF. Youth problem behaviors 8 years after implementing the Communities That Care prevention system. A community-randomized trial. JAMA Pediatrics. 2014; 168:122–129. [PubMed: 24322060]

Hawkins JD, Oesterle S, Brown EC, Arthur MW, Abbott RD, Fagan AA, Catalano RF. Results of a type 2 translational research trial to prevent adolescent drug use and delinquency: A test of Communities That Care. Archives of Pediatrics and Adolescent Medicine. 2009; 163:789–798. [PubMed: 19736331]

Hawkins JD, Oesterle S, Brown EC, Monahan KC, Abbott RD, Arthur MW, Catalano RF. Sustained decreases in risk exposure and youth problem behaviors after installation of the Communities That

Care prevention system in a randomized trial. Archives of Pediatrics and Adolescent Medicine. 2012; 166:141–148. [PubMed: 21969362]

Henderson AR. Assessing test accuracy and its clinical consequences - a primer for receiver operating characteristic curve analysis. Annals of Clinical Biochemistry. 1993; 30:521–539. [PubMed: 8304720]

Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, … Wang PS. The epidemiology of major depressive disorder - Results from the National Comorbidity Survey Replication (NCS-R). Journal of the American Medical Association. 2003; 289:3095–3105. [PubMed: 12813115]

Kroenke K, Spitzer RL, Williams JBW. The PHQ-9 - Validity of a brief depression severity measure. Journal of General Internal Medicine. 2001; 16:606–613. [PubMed: 11556941]

Monahan KC, Oesterle S, Rhew I, Hawkins JD. The relation between risk and protective factors for problem behaviors and depressive symptoms, antisocial behavior, and alcohol use in adolescence. Journal of Community Psychology. 2014; 42:621–638. [PubMed: 26985122]

Revelle, W. psych: Procedures for psychological, psychometric, and personality research. Evanston, IL: Northwestern University; 2014.

Rhew IC, Simpson K, Tracy M, Lymp J, McCauley E, Tsuang D, Stoep AV. Criterion validity of the Short Mood and Feelings Questionnaire and one- and two-item depression screens in young adolescents. Child and Adolescent Psychiatry and Mental Health. 2010; 4:8. [PubMed: 20181135]

Richardson LP, McCauley E, Grossman DC, McCarty CA, Richards J, Russo JE, … Katon W. Evaluation of the Patient Health Questionnaire-9 Item for detecting major depression among adolescents. Pediatrics. 2010; 126:1117–1123. [PubMed: 21041282]

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. pROC: An open-source package for R and S plus to analyze and compare ROC curves. BMC Bioinformatics. 2011; 12:77. [PubMed: 21414208]

Rosseel Y. lavaan: An R package for structural equation modeling. Journal of Statistical Software. 2012; 48:1–36.

Sharp C, Goodyer IM, Croudace TJ. The Short Mood and Feelings Questionnaire (SMFQ): A unidimensional item response theory and categorical data factor analysis of self-report ratings from a community sample of 7- through 11-year-old children. Journal of Abnormal Child Psychology. 2006; 34:379–391. [PubMed: 16649000]

Spoth R, Redmond C, Shin C, Greenberg M, Clair S, Feinberg M. Substance-use outcomes at 18 months past baseline: The PROSPER Community-University Partnership Trial. American Journal of Preventive Medicine. 2007; 32:395–402. [PubMed: 17478265]
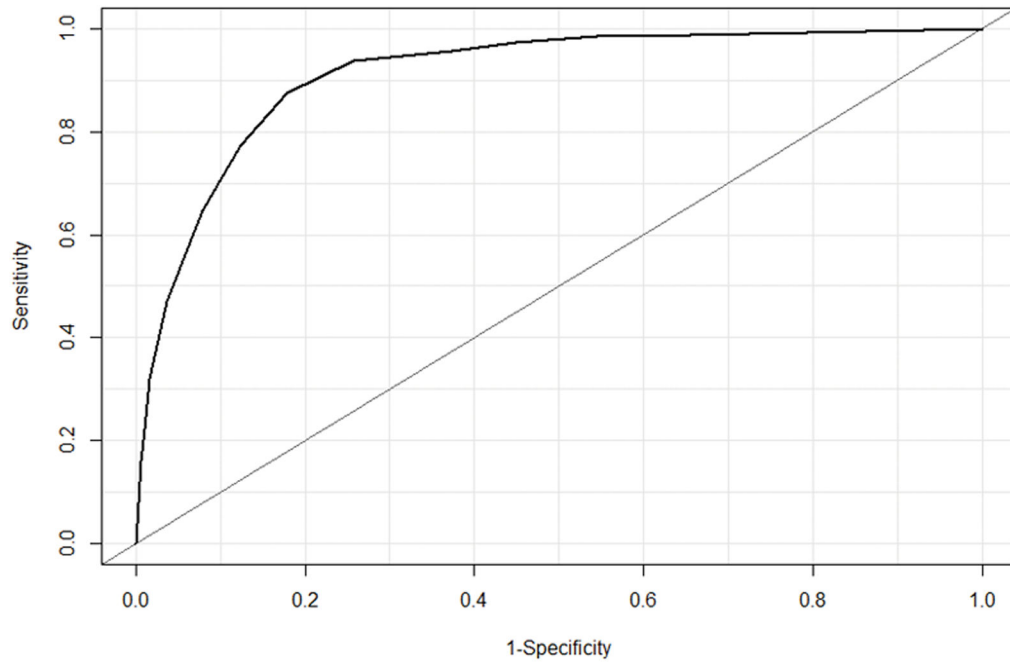
Figure 1.
Receiver operating characteristic curve for the CTC Brief Depression Scale against major depressive disorder diagnosis.