Improving Suicide Screening at the Cleveland Clinic through Electronic Self-Reports: PHQ-9 and the Columbia-Suicide Severity Rating Scale (C-SSRS)

Irene L. Katzan[1], M.D.; Adele C. Viguera[1], M.D., M.P.H; Taylor Burke[2], B.A.; Jacqueline Buchanan[2], A.B.; Kelly Posner[2], Ph.D.
[1]Cleveland Clinic [2]Columbia University Medical Center

BACKGROUND

- The Joint Commission has indicated suicide screening a National Patient Safety Goal (The Joint Commission, 2010). As a result, emergency rooms, clinics, and primary care divisions within hospitals are increasingly implementing suicide screening protocols.
- Utilizing self-report screening measures has been indicated as a feasible, low-burden way to systematically screen and monitor individuals for suicide risk (Horowitz et al., 2009).
- Many primary care physicians and hospitals utilize the PHQ-9, a 9-item depression screener, which includes one item about self-harm (item 9): ***"thoughts that you would be better off dead or of hurting yourself."*** Although this single, binary question has been widely used as a suicide risk screen, it does not specifically assess suicidal ideation or behavior.
- In contrast, the Columbia-Suicide Severity Rating Scale (C-SSRS) is a validated suicide risk assessment tool that evaluates the full range of suicidal ideation and behaviors (Posner et al., 2011). This tool also provides a summary measure (positive or negative screen) of suicide risk, as well as categorical outcomes for ideation and behavior.

PURPOSE

- Our aim was to examine the psychometric properties of the PHQ-9 suicide item - specifically the probability of oversampling due to imprecision of the question stem. Sensitivity, specificity and negative/positive predictive values of the PHQ-9 item were evaluated using an electronic self-report screening version of the Columbia-Suicide Severity Rating Scale (C-SSRS) as the gold standard.

METHODS

- As part of a project to develop an electronic suicide screening and safety algorithm for the Cleveland Clinic (including psychiatry, psychology, neurology, neurosurgery, and rehabilitative medicine), the PHQ-9 and a self-report screening version of the C-SSRS were collected electronically from adult ambulatory psychiatry patients from December 14, 2011 through April 4, 2012.
- 1,461 patients responded to item-9 of the PHQ-9 and the C-SSRS. 1,457 patients responded to the entire PHQ-9 and the C-SSRS.
- 1,572 patients completed the C-SSRS. For all 1,572 patients, average age was 43.8 years (SD = 14.9 years), with 62.8% female and 75.5% Caucasian (Table 1).
- The first C-SSRS and PHQ-9 scores obtained from each patient during the assessment period were used for all analyses.
- PHQ-9 item 9 sensitivity, specificity, and positive and negative predictive values were calculated.

RESULTS

- The observed point prevalence of suicidal ideation, behavior, or both was 6.2% (98/1572; CI: 5.0-7.4%) on the C-SSRS and was 23.8% (347/1461; CI: 21.6-25.9%) on PHQ-9 item 9, a nearly 4-fold increase in positive screens.
- PHQ-9 Item 9 yielded a sensitivity of 91.8% (78/85; CI 85.9-97.6%).
- 7 individuals (<1%) with a positive C-SSRS did not endorse item 9, leading to a negative predictive value of 99.4% (1107/1114; CI: 99.0%-99.8%).
- Many who endorsed item 9 did not screen positive for suicidal ideation or behavior on the C-SSRS Screen, resulting in a specificity of only 80.5% (1,107/1,376; CI: 78.4-82.5%) and very low positive predictive value of 22.5% (78/347; CI: 20.3-24.6%).
- As the PHQ-9 total score increased, the frequency of a positive C-SSRS screen also increased (Table 1).
- Summary statistics for patients who tested positive or negative for suicide risk on the C-SSRS screen are provided in Table 1. The average C-SSRS positive PHQ-9 score for this group was 19.0 (SD = 6.2) and the average C-SSRS negative PHQ-9 score was 9.2 (SD = 7.0).

	Positive Columbia	Negative Columbia	Total	P-value
Age				
N	98	1,474	1,572	0.0579*
Mean (SD)	41.2 (13.9)	43.9 (15.0)	43.8 (14.9)	
Median (IQR)	44.0 (27.8-52.0)	44.0 (31.0-55.0)	44.0 (31.0-55.0)	
Gender				
Male	43 (43.9%)	541 (36.7%)	584 (37.2%)	0.1884**
Female	55 (56.1%)	933 (63.3%)	988 (62.8%)	
PHQ-9 Score				
N	85	1,372	1,457	< 0.0001*
Mean (SD)	19.0 (6.2)	9.2 (7.0)	9.7 (7.3)	< 0.0001**
Median (IQR)	21.0 (15.0-24.0)	8.0 (3.0-14.0)	8.0 (4.0-15.0)	
0-4 (None)	3 (3.5%)	434 (31.6%)	437 (30.0%)	
5-9 (Mild)	4 (4.7%)	364 (26.5%)	368 (25.3%)	
10-14 (Moderate)	14 (16.5%)	253 (18.4%)	267 (18.3%)	
15-20 (Moderate-Severe)	17 (20.0%)	171 (12.5%)	188 (12.9%)	
20-27 (Severe)	47 (55.3%)	150 (10.9%)	197 (13.5%)	
Item 9 of PHQ-9				
0	7 (7.1%)	1,107 (75.1%)	1,114 (70.9%)	0.0005***
1	30 (30.6%)	197 (13.4%)	227 (14.4%)	
2	18 (18.4%)	40 (2.7%)	58 (3.7%)	
3	30 (30.6%)	32 (2.2%)	62 (3.9%)	
Did not Answer	13 (13.3%)	98 (6.6%)	111 (7.1%)	
Diagnosis				
Major Depression	37 (37.8%)	415 (28.2%)	452 (28.8%)	0.0070***
Anxiety	3 (3.1%)	159 (10.8%)	162 (10.3%)	
Bipolar	11 (11.2%)	133 (9.0%)	144 (9.2%)	
Psychosis	8 (8.2%)	66 (4.5%)	74 (4.7%)	
Other	6 (6.1%)	184 (12.5%)	190 (12.1%)	
Unknown	33 (33.7%)	517 (35.1%)	550 (35.0%)	

Table 1. Baseline summary statistics for entire population and separated by patients who had a positive screen and a negative screen on the Columbia. P-values for comparing patients with a positive C-SSRS Screen to negative C-SSRS Screen are also provided * calculated from a two-sample t-test for the difference in means; ** calculated from a Chi-square test of independence; *** calculated using Monte Carlo simulation

	Negative Columbia Screen	Positive Columbia Screen	Totals
Did not Endorse Item 9 on PHQ-9	1,107	7	1,114
Endorsed Item 9 on PHQ-9	269	78	347
Totals	1,376	85	1,461

Table 2. Counts of patients who endorsed/did not endorse Item 9 of the PHQ-9 and had a positive/negative C-SSRS.

	Estimate	95% CI Lower Bound	95% CI Upper Bound
Prevalence *	6.2%	5.0%	7.4%
Sensitivity **	91.8%	85.9%	97.6%
Specificity **	80.5%	78.4%	82.5%
Positive Predictive Value **	22.5%	20.3%	24.6%
Negative Predictive Value **	99.4%	99.0%	99.8%

Table 3. Prevalence of positive C-SSRS screen and sensitivity and specificity of Item 9 of the PHQ-9, along with 95% confidence intervals.* Based on all 1,568 patients who completed the C-SSRS Screen. ** Based on 1,461 patients who completed both Item 9 and the C-SSRS Screen.

DISCUSSION

- Relying solely on PHQ-9 item 9 as a suicide screen may result in a significant rate of false positives, given its modest specificity (80.5%) and low positive predictive value (22.5%). This finding is consistent with structured clinical interview data collected from cancer patients an average of 10 days after screening positive on the PHQ-9 indicating that only 1/3 of patients exhibited clear thoughts of killing themselves (Walker et al., 2011).
- False positives may be explained in part by the conflation of suicidal and non-suicidal self-injurious thoughts. A substantial proportion of those who exhibit NSSI thoughts and behaviors do not think about suicide or engage in suicidal behaviors (Jacobson et al., 2008). However, the presence of NSSI is a significant risk factor for future suicidal behaviors (Hamza et al., 2012).
- Notably, the timeframe of the PHQ-9 (2 weeks) is shorter than the ideation (1 month) and behavior (3 months) timeframes of the C-SSRS. That the high rate of false positives occurred in spite of a much shorter timeframe suggests that results may be understated.
- It will be important for hospital-wide screening efforts to reduce false positives to ensure the feasibility of implementing safety procedures for those who are at true risk for suicide.
- In capturing a wide range of ideation and behavior types (preparatory behaviors, aborted and interrupted suicide attempts, and actual suicide attempts), the C-SSRS may perform better than the PHQ-9 in capturing thoughts/behaviors that are usually not queried, thereby reducing false negatives.
- The PHQ-9 had 7 false negative cases; researchers should continue to refine suicide screening tools to minimize false negatives, as identifying individuals at risk is of critical importance.

LIMITATIONS

- Due to the fact that this study was completed within an outpatient psychiatric population with a relatively high baseline prevalence of suicidal ideation and behavior, our results may not be generalizable to the community. Therefore, the psychometric properties of the PHQ-9 (as compared to the C-SSRS) should be further examined among a diverse community sample.
- The use of cross-sectional data prevented our ability to follow up with patients to calculate actual risk of engaging in future suicidal behavior (as opposed to theoretical risk based on self-reported scores).
- Further, the disparate timeframes employed by the PHQ-9 and the C-SSRS preclude our ability to discern the origin of the highly divergent rates of risk detection. In order to determine the clinical significance of each of the question stems, future studies should employ parallel timeframes.