Received: 21 January 2021 | Revised: 12 May 2021 | Accepted: 14 May 2021

DOI: 10.1111/sltb.12791

Suicide and Life-Threatening BEHAVIOR

ORIGINAL ARTICLE

Examining suicide assessment measures for research use: Using item response theory to optimize psychometric assessment for research on suicidal ideation in major depressive disorder

William V. McCall MD, MS[1] | Ben Porter PhD[2] | Ashley R. Pate MS[3] | Courtney J. Bolstad MS[3] | Christopher W. Drapeau PhD[4,5] | Andrew D. Krystal MD, MS[6] | Ruth M. Benca MD, PhD[7] | Meredith E. Rumble PhD[8] | Michael R. Nadorff PhD[3,9]

[1]Department of Psychiatry and Health Behavior, Medical College of Georgia, Augusta University, Augusta, Georgia, USA

[2]Social Science Research Center, Mississippi State University, Mississippi State, Mississippi, USA

[3]Department of Psychology, Mississippi State University, Mississippi State, Mississippi, USA

[4]Department of Health Policy and Management, Indiana University Richard M. Fairbanks School of Public Health, Indianapolis, Indiana, USA

[5]Division of Mental Health and Addiction, Indiana Family and Social Services Administration, Indianapolis, Indiana, USA

[6]University of California San Francisco, San Francisco, California, USA

[7]Department of Psychiatry and Human Behavior, University of California, Irvine, California, USA

[8]Department of Psychiatry, University of Wisconsin, Madison, Wisconsin, USA

[9]Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, Texas, USA

Correspondence
William V. McCall, Department of Psychiatry and Health Behavior, Medical College of Georgia, Augusta University, 997 St Sebastian Way, Augusta, GA 30912, USA.
Email: wmccall@augusta.edu

Funding information
National Institute of Mental Health, Grant/Award Number: MH095776, MH095778 and MH095780

Abstract

Introduction: Progress reducing suicide death will require randomized clinical trials (RCTs) specifically targeting suicide risk. Even large RCTs may not stipulate suicide death as the primary outcome, as suicide death is relatively uncommon. Therefore, RCTs may need to specify suicidal ideation as a proxy indicator of risk. There is no consensus on the best tool for measuring suicidal ideation within RCTs. We contrasted the psychometric performance of three suicidal ideation measures to address this need.

Methods: We applied item response theory to the Beck Scale for Suicide Ideation (BSSI), the Columbia-Suicide Severity Rating Scale (C-SSRS), and the suicide item of the Hamilton Rating Scale for Depression (HRSD) for 101 outpatients with depression and suicidal ideation participating in a RCT with suicidal ideation as the primary outcome.

Results: All measures of suicidal ideation were equally able to detect low and very high levels of suicidal ideation.

Conclusions: The choice of the specific measure of suicidal ideation in a clinical trial may be dictated by time and financial resources versus the need for granularity in the interpretation of the scores.

INTRODUCTION

Suicide is the 10th leading cause of death in the United States and accounted for 48,344 deaths in 2018 (Drapeau & McIntosh, 2020). Unfortunately, despite decades of research little progress has been made on preventing suicide, as the rate of suicide has been increasing each year since 1999 (Drapeau & McIntosh, 2020). Indeed, a recent meta-analysis of 50 years of suicide research indicated that current methods of predicting suicide were comparable to chance, a prediction rate that has not improved in the last 50 years (Drapeau & McIntosh, 2020; Franklin et al., 2017). Suicidal ideation is one of several clinical predictors for suicide, and several measures for assessing suicidal ideation and suicide risk are in current use, but a clear gold standard has not been established (Batterham et al., 2015; Brown, 2002; Gutierrez et al., 2020). Thus, there is a great need for additional studies examining how to best choose among measures or clinical tools that assess suicidal ideation.

Current measures of suicidal ideation and behaviors

There are a wide range of assessment tools for measuring suicidal thoughts and behaviors. These assessments vary greatly in length, ease of administration (e.g., some require trained assessors; Jobes, 2006), or copyright protection (Reynolds, 1991). Additionally, certain measures (e.g., Kroenke et al., 2001) have single items assessing for passive suicidal ideation (Kroenke et al., 2001), which may identify many individuals endorsing thoughts of suicide without intent to exhibit future suicidal behavior. Alternatively, other measures assess only severe symptoms and thus may misidentify individuals at elevated but not extreme risk for suicide (Hom et al., 2016).

Self-reported measures are less costly to administer, and there tends to be a high degree of agreement between self-report and interviewer-administered assessments. There is some evidence that individuals are more likely to report suicidal ideation on self-report (Beck et al., 1988; Kaplan et al., 1994). Especially with large studies or studies where suicidal ideation is a secondary outcome, measures may need to be brief to reduce participant burden, but longer and more comprehensive measures may be preferred for studies where suicide risk is the primary focus.

However, test attributes and optimal levels of sensitivity can change depending on the situation. For example, a copy written test may be appropriate for use in a small clinic but is not feasible to use in large population studies (Batterham et al., 2015). Similarly, assessing low levels of symptoms may be appropriate for identifying a pool of at-risk participants but would likely result in too many false-positive cases for an intensive intervention. For intervention clinical trials, a measure is needed that (1) will accurately assess suicidal ideation or suicide risk, (2) is sensitive to change, (3) will differentiate between a therapeutic intervention versus a placebo/sham, and (4) have enough granularity such that incremental reductions of suicidal ideation or risk can be translated into clinically meaningful gradations of suicide risk (i.e., presence or absence of a plan, and presence or absence of intent to die by suicide). Additionally, measures that reduce participant burden and cost to the study (to either purchase or administer) are ideal.

Lacking a gold-standard assessment for research

One of the most notable limitations of the suicide literature, particularly the suicide treatment literature, is the lack of consensus for a gold-standard assessment for suicidal ideation in clinical trials. The last major undertaking for a review of various suicide assessments took place nearly two decades ago (Brown, 2002). In this review, Brown concluded the necessity of both self-report and interviewer-administrated measures. Although this review established some assessment tools as preferred due to their predictive validity (e.g., Beck Scale for Suicide Ideation [BSSI] and Beck Hopelessness Scale [BHS]), there was not one agreed upon assessment of choice due to the complexity of suicidal thoughts and behaviors and the needs of different settings (clinical versus research) and different populations, indicating the need for additional research on this topic (Brown, 2002). In contrast, nearly every antidepressant clinical trial utilizes the Hamilton Rating Scale for Depression (HRSD; Millner et al., 2015). However, a review of suicidal ideation measures did not find any that met desired properties for population-level research or discusses preferred measures for clinical trials (Batterham et al., 2015).

Since Brown's review, other suicide assessments have been developed, the most notable of which is the Columbia-Suicide Severity Rating Scale (C-SSRS; Posner et al., 2011) becoming one of the most widely used suicide assessments because it has established predictive validity (Gipson et al., 2015). The FDA supported the use of the C-SSRS for suicide risk surveillance tool in clinical trials (Center for Drug Evaluation and Research, 2012). Although the Veteran Health Administration screens for suicide risk with a single item from the PHQ-9 (Kroenke et al., 2001), positive screens are further assessed with the C-SSRS (Katz et al., 2020). However, critics suggest the C-SSRS may lead to over-endorsement of suicidal ideation (Giddens et al., 2014), and the semi-structured interview adds significant burden when compared to a self-report measure.

Part of the difficulty in optimizing suicidal ideation assessment is because suicide is a low-probability behavior

in high-risk groups, a property that is exacerbated when examining the general population. Therefore, only a few measures (including the BSSI, BHS, and C-SSRS) have predictive validity through associations with future suicide attempts and deaths (Brown, 2002; Katz et al., 2020; Lindh et al., 2019). Although many suicide assessment measures are highly correlated, others have small and negligible correlations (Gutierrez et al., 2019), either indicating the measures have either poor validity or poor coverage of suicide risk factors. Interestingly, single item measures have been shown to have high correlations with longer more exhaustive measures of suicide risk (Brown, 2002; Ducher & Dalery, 2008).

A second limitation in identifying a gold standard for suicide assessment in clinical trials is the scarcity of studies comparing the simultaneous performance of multiple measures. One early example of this comes from Cochrane-Brink and Sakinofsky who compared several suicide and depression measures to see which best predicted admission to the hospital for suicide concerns (Cochrane-Brink et al., 2000). They found the High-Risk Construct Scale best predicted suicide admission, followed by the BSSI (Cochrane-Brink et al., 2000). Although this is an interesting finding, it does not tell us anything about the sensitivity to change in suicidal ideation over time. Lindh and colleagues built upon this work through comparing the suicide intent scale, suicide assessment scale, Karolinska interpersonal violence scale, and the C-SSRS suicidal ideation scale in predicting suicide attempts and death by suicide (Lindh et al., 2019). This is a notable advance as the outcomes are suicidal behavior and not clinical judgment regarding suicidal risk. However, they found that measures that predicted multiple suicide attempts failed to predict death by suicide and vice versa, and the specificity was low, making it so no one measure emerged as the preferred measure. Recently, Gutierrez and colleagues compared the predictive validity of C-SSRS, BSSI, Self-Harm Behavior Questionnaire (SHBQ; Gutierrez et al., 2001), and the SBQ-R within a large group of military service members identified as being at increased risk for suicide (Gutierrez et al., 2020). Although certain measures had statistically significant associations with subsequent suicide attempts and future ideation, the performance of all measures was practically similar such that no single measure or combination of measures was identified as superior at predicting attempts and ideation.

Statement of the problem

There is no accepted gold-standard measure for suicidal ideation as the primary outcome within clinical trials. There also are no measures that have proven to meet the needs of (1) accurately assessing suicidal ideation, (2) being sensitive to changes in ideation, (3) differentiating between a therapeutic intervention versus a placebo/sham, and (4) having enough granularity such that incremental reductions of suicidal ideation or risk can be translated into clinically meaningful gradations of suicide risk. However, item response theory can be used to determine how well measures perform over the continuum of suicidal ideation. Given that recent research suggests that no measure preforms universally superior to the others (Gutierrez et al., 2020; Lindh et al., 2019), shorter measures with lower administration burden should be compared to longer semi-structured interviews such as the C-SSRS.

Current study

The present study sought to fill this gap in the literature by comparing suicidal ideation assessment tools within a clinical trial. The C-SSRS, BSSI, and the single suicide item from the HRSD were compared using item response theory over the course of a clinical trial. This study is notable as it is the first to examine which measure of suicidal ideation is preferable for clinical research instead of focusing on variables related to clinical practice.

MATERIALS AND METHODS

Overview

This is a secondary analysis of data from the parent study "Reducing suicidal ideation through insomnia treatment (REST-IT)." The REST-IT study was an 8-week, 3-site, double-blind, placebo-controlled, parallel-group randomized controlled trial of zolpidem controlled release (CR) hypnotic therapy versus placebo, in conjunction with open-label selective serotonin reuptake inhibitor (SSRI). At baseline, participants were medication-free 18- to 65-year-olds, with major depressive disorder, insomnia, and suicidal ideation. Suicidal ideation was measured with the BSSI, the C-SSRS suicidal ideation scale, and the suicide item within the HRSD. Inclusion in the study required a BSSI score ≥ 3, and all participants provided written informed consent after receiving a complete description of the study. Participants received open-label SSRI and randomized bedtime study drug together after the baseline visit and were followed up at 1, 2, 4, 6, and 8 weeks. Suicidal ideation was measured at baseline and every subsequent visit. The data from the two randomized groups were combined for the purposes of this secondary analysis. Additional information about the parent study is available elsewhere (McCall et al., 2019).

Measures

The BSSI is a self-rated scale consisting of 19 items that evaluate suicidal desire and planning (Beck et al., 1997; Beck et al., 1999; Beck et al., 1979) Each item is rated on a 3-point scale from 0 to 2 for a maximum score of 38; a lower score indicates less severe suicidal ideation. A BSSI ≥3 is a significant predictor for suicide death over a period of up to 20 years (hazard ratio = 6.6; Brown et al., 2000). The C-SSRS is an observer-rated scale (Oquendo et al., 2003; Posner et al., 2007, 2011). The suicidal ideation "intensity" dimension of the C-SSRS is scored 0–5, with five representing suicidal ideation with a plan and intent. The 24-item HRSD includes a single item assessing suicide which is scaled 0–4, with "4" indicating a serious suicide attempt (Hamilton, 1960). The HRSD was administered by trained research staff, while the C-SSRS was administered by doctoral level investigators.

The BSSI scores were subdivided into five different levels based on the observed frequencies in the current study. Approximately 50% of observations had responses less than six which corresponds roughly to the lowest categories of risk for the other two measures. Three additional groups were created with a range of five points (i.e., 6–10, 11–15, 16–20). Only a small proportion of participants had a BSSI score above 20 (3.2%) and so this cut point was used for all values above this point.

Given that the goal for most suicide interventions is to eliminate risk of suicide rather than lessening its likelihood, all three measures were also examined as dichotomous variables where "0" indicated a participant endorsed the lowest level score on each measure and "1" indicated a participant provided a response above this level.

Analysis plan

The measurements from both treatment groups were combined for the purposes of this secondary examination of the psychometric properties of the three suicidal ideation measures, using item response theory. Item response theory assumes there is a continuous unobserved underlying dimension (i.e., "suicidal ideation") that is indicated by responses to given items and measures. Scores on the three measures used to estimate this underlying dimension and in turn identify where the measures and items distinguish individuals' levels of that dimension.

In the current application, a graded response model was used to estimate nominal responses to measures (Samejima, 2016). This model estimates a single slope and $k - 1$ thresholds for each measure where k indicates the number of response options. Separately, a two-parameter logistic model was used to estimate the association between dichotomous measures and suicidal ideation (similar to the graded response model but only one threshold is estimated). In both these models, the slope indicates the strength of association between the measure and suicidal ideation, whereas the thresholds indicate the *Z*-score of an individual who has an equal probability of selecting a given response option or higher. With only three measures, the data were assumed to be unidimensional. All analyses were performed using the IRT procedures within SAS software.

An initial graded response model was conducted that included both baseline and all available follow-up assessments. Bayesian information criteria (BIC) from this initial model were compared to those of three different models in which the parameters of the BSSI, C-SSRS suicidal ideation scale, and the HRSD item were allowed to vary between baseline and follow-up and the remaining two variables acted as anchors. Smaller BIC indicated a superior model fit.

TABLE 1 Demographic and baseline characteristics of 101 adults with suicidal ideation

Characteristic	Frequency (%)
Age (years, *M* [SD])	40 (13.2)
Sex	
Male	37 (36.6%)
Female	64 (63.4%)
Race/ethnicity	
Caucasian	62 (61.4%)
African American	27 (36.7%)
Hispanic	5 (5.0%)
Other	7 (6.9%)
Baseline BSSI, *M* (SD)	12.0 (5.3)
Baseline C-SSRS suicidal ideation scale, *M* (SD)	1.6 (1.0)
Baseline HRSD suicidal ideation item, *M* (SD)	1.7 (0.7)

Abbreviations: BSSI, Beck Scale for Suicide Ideation; C-SSRS, Columbia-Suicide Severity Rating Scale; HRSD, Hamilton Rating Scale for Depression.

RESULTS

The parent study sample consisted of 103 participants who completed baseline assessments, but two participants were excluded as they were randomized but did not complete follow-up assessments. Therefore, 101 baseline observations and 455 follow-up assessments were used in analysis. Overall, the study population was middle aged, with a majority of women and Caucasians, but with good representation of minority racial/ethnic groups (Table 1). Scores on the three measures at baseline indicated notable suicidal ideation, with the average score on the BSSI being more than twice the clinical cutoff at baseline (Sokero et al., 2003).

The three suicidal ideation measures were highly correlated (rs > 0.60, ps < 0.001). Additionally, when dichotomized to reflect an absence of suicidal ideation, 74% of observations had the same indication on all three measures. The most common disagreement between dichotomized measures occurred when the BSSI indicated suicidal ideation but the other two measures did not (12% of cases). Overall, participants reported more severe suicidal ideation at baseline than at follow-up assessments (Table 2). At baseline, low and high values (i.e., indicating low or high levels of suicidal ideation) were relatively uncommon, whereas at follow-up, only high values were uncommon.

A graded response model was conducted on the three measures using baseline and follow-up assessments combined. The decision to select a single factor solution was validated by the scree plot and eigenvalues (2.52, 0.31, 0.17). BIC for this model was 3426.8. The model in which HRSD was allowed to vary between baseline and follow-up assessments indicated slightly superior fit (BIC = 3424.6). However, the BIC of the other two models indicated that this relaxation of parameters did not improve fit (C-SSR BIC = 3444.3; BSSI BIC = 3438.1). The parameters from the model relaxing equality among HRSD are displayed in Table 3.

Graded response parameters indicate that HRSD was more strongly associated with suicidal ideation (i.e., the other two measures) at follow-up than at baseline. Overall, the HRSD item provided the most information at follow-up, but C-SSRS provided the most information at baseline (see Figure 1). However, the range of suicidal ideation at which the measures performed best (i.e., had the highest information) was relatively similar, ranging from approximately 0–3 standard deviations above the mean. This indicates that none of these measures distinguishes between individuals with low suicidal ideation versus individuals with very low suicidal ideation.

A two-parameter logistic model was constructed using the three dichotomized variables. As before, a single factor was suggested (eigenvalues 2.74, 0.21, 0.05). Nonconvergence prevented identifying parameter differences on the baseline and follow-up assessments. Slopes were generally high, but were lowest for the BSSI, followed by the HRSD item and then the C-SSRS (2.83, 7.72, 7.91, respectively). The thresholds were all close to zero (BSSI, −0.41; C-SSRS, −0.03; and HRSD, −0.05). As would be expected from the larger slopes, item information was highest for the C-SSRS and HRSD (Figure 2).

DISCUSSION

The current investigation is one of the first to directly compare the psychometric performance of multiple measures of suicidal ideation *in the context of a clinical trial* (McCall et al., 2019). The use of item response theory allowed for a detailed description of the measures' relative performance.

TABLE 2 Frequencies of response options at baseline and follow-up

Measure	Value	Baseline (%)	Follow-up (%)
Beck Scale for Suicide Ideation	0–5	9 (8.9)	272 (59.8)
	6–10	33 (32.7)	89 (19.6)
	11–15	38 (37.6)	60 (13.2)
	16–20	14 (13.9)	20 (4.4)
	20–33	7 (6.9)	14 (3.0)
Columbia-Suicide Severity Rating Scale: suicidal ideation subscale	Lack of suicidal ideation (score 0)	12 (11.9)	254 (55.8)
	Wish to be dead (score 1)	40 (39.6)	124 (27.3)
	Non-specific active suicidal thoughts (score 2)	23 (22.8)	42 (9.2)
	Active suicidal ideation with any methods (not plan) without intent to act (score 3)	25 (24.8)	30 (6.6)
	Active suicidal ideation with some intent to act, without a specific plan (score 4)	1 (1.0)	3 (0.7)
	Active suicidal ideation with specific plan and intent (score 5)	0 (0)	0 (0)
Hamilton Rating Scale for Depression: suicidal ideation/behavior item	Absent	8 (7.9)	255 (56.0)
	Feels life is not worth living	24 (23.8)	111 (24.4)
	Wishes he/she were dead or any thoughts of possible death to self	63 (62.4)	78 (17.1)
	Ideas or gestures of suicide	6 (5.9)	11 (2.4)
	Attempts at suicide	0 (0.0)	0 (0.0)

TABLE 3 Graded response model parameters

Measure	Baseline slope	Follow-up slope	Threshold criteria	Baseline threshold	Follow-up threshold
Beck Scale for Suicide Ideation	2.28	—	0–5	N/A	—
			6–10	0.04	—
			11–15	0.78	—
			16–20	1.62	—
			20–33	2.21	—
Columbia-Suicide Severity Rating Scale: suicidal ideation subscale	3.57	—	Lack of suicidal ideation	N/A	—
			Wish to be dead	−0.02	—
			Non-specific active suicidal thoughts	0.89	—
			Active suicidal ideation with any methods (not plan) without intent to act	1.39	—
			Active suicidal ideation with some intent to act, without a specific plan	2.55	—
Hamilton Rating Scale for Depression: suicidal ideation/behavior item	1.99	5.36	Absent	N/A	N/A
			Feels life is not worth living	−0.89	0.01
			Wishes he/she were dead or any thoughts of possible death to self	0.22	0.75
			Ideas or gestures of suicide	2.50	1.84

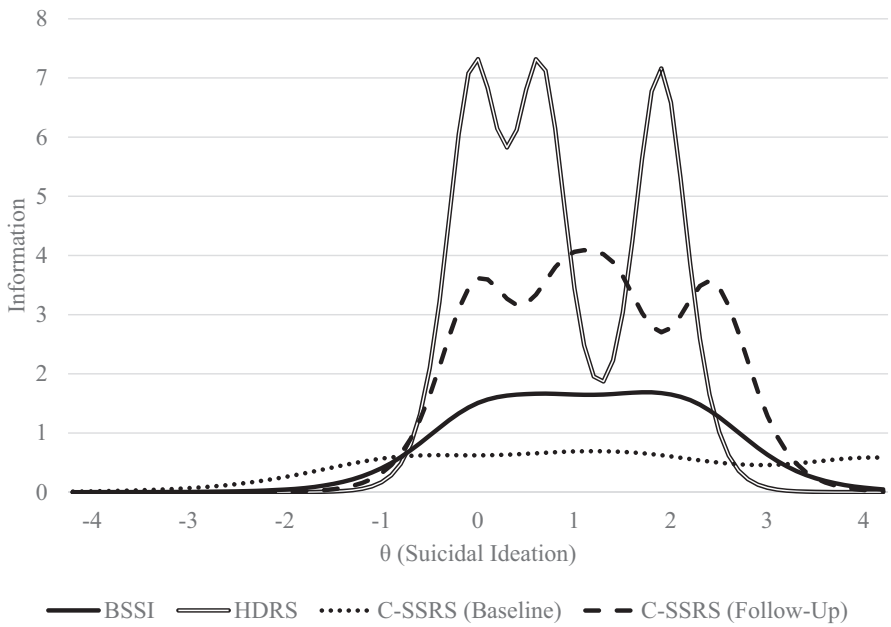

FIGURE 1 Item information curves from categorical assessment. BSSI, Beck Scale for Suicide Ideation; C-SSRS, Columbia-Suicide Severity Rating Scale; HRSD, Hamilton Rating Scale for Depression

While the slopes and information scores were not statistically compared, the C-SSRS and HRSD had the highest slopes and information of the three measures examined. However, this may be due to the arbitrary cut points imposed on the BSSI that may not match the qualitative cut points of the other two measures. It is possible that had different cut points been chosen, the BSSI would have performed better.

Additionally, the three measures assess suicidal ideation across a similar range of latent values, with the highest information occurring approximately between 0 and 2 standard deviations above the mean of the current sample. This

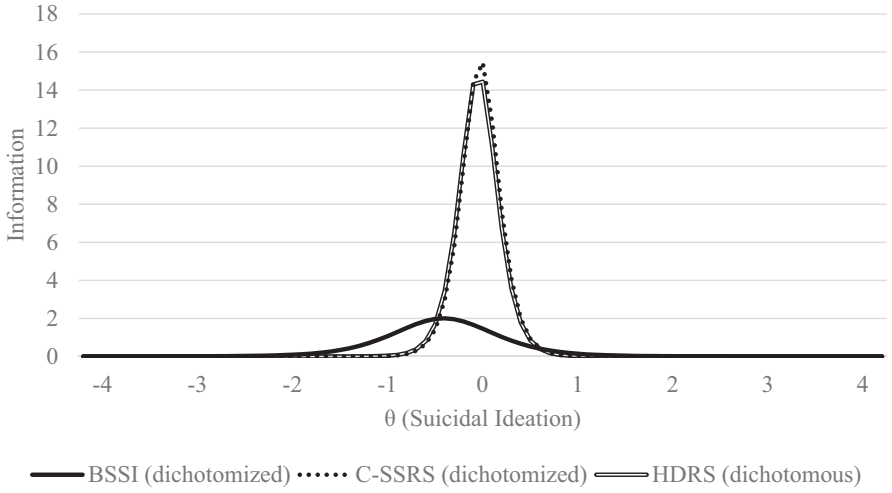

FIGURE 2 Item information curves from dichotomous assessment. BSSI, Beck Scale for Suicide Ideation; C-SSRS, Columbia-Suicide Severity Rating Scale; HRSD, Hamilton Rating Scale for Depression

suggests that none of the scales is particularly better than the others at distinguishing between low versus very low suicidal ideation. However, given the practical needs for detecting suicide risk, good performance in the low to very low range would be desirable and hence reveals a limitation of all three scales. Comparisons of the psychometric properties of depression scales have concluded that self-report scales may be less sensitive to change as compared against observer-rated depression scales (Castrogiovanni et al., 1989; Hallam et al., 2009; Panel, 1993). Similarly, this may explain why the BSSI failed to distinguish between randomized active treatment versus placebo in the parent study, while the C-SSRS suicidal ideation scale did differentiate between treatments (McCall et al., 2019).

Limitations

This secondary analysis had several limitations and strengths. As a statistical procedure, item response theory works best with many observations of which the current analysis had near the minimum recommended sample size. Models that failed to converge would likely have converged with a larger sample size. However, almost all models converged, potentially due to the simplicity of the model examined. Additionally, the current analysis did not control for the nested nature of the data due to loss in analytic power. This is expected to increase the agreement between measures to the extent that individuals responded consistently across time; however, both experimental and control groups experienced significant declines in symptoms across time (McCall et al., 2019), limiting the impact of ignoring nesting. Finally, the results of the item response theory rely heavily on the measures/items and sample used to estimate the model. As such, were additional measures of suicidal ideation included or analyses conducted among a different sample, the results may have differed. Therefore, the results of this study should be taken as preliminary suggestions regarding suicidal ideation metrics in the context of randomized clinical trials and future studies should conduct similar analyses to validate or refute the findings herein.

It is uncertain whether these findings inform the choice of which suicidal ideation metric to pick in a routine *clinical setting*. Also, this study only examined suicidal ideation and cannot comment on psychometric performance concerning suicidal behavior, as there was no suicidal behavior in the course of the parent study. An additional limitation was the need for an inclusion criterion regarding the minimum amount of suicidal ideation required for a participant to enter the study. The parent study chose a BSSI score ≥ 3 as that criterion, which is indeed a low level of suicidal ideation. However, no such requirements were made for the C-SSRS or HRSD. It is possible that the BSSI entry requirement affected our analysis.

The study has some strengths that bear mentioning. Multiple high-quality indicators of suicidal ideation were collected at each observation by clinical staff in the context of a research study, which represents a notable advance in this literature. Also, the sample used in the current study is comparable to that of other treatment studies, and as such, the results can likely be generalized to other clinical research studies.

Pragmatically looking at the performance of the three measures, there was strong agreement between the measures as measured by correlation and correspondence of dichotomous indicators of the presence/absence of suicidal ideation. However, there are notable differences between the measures. First, the values obtained from the C-SSRS suicidal ideation scale have the most clinically intuitive and relevant interpretation, as the scores move in logical progression from no suicidal ideation, to passive thoughts of death, to non-specific thoughts of suicide, to specific means, to intent, and finally to specific plan. In comparison, the total scores from the BSSI cannot be interpreted

in a similar manner, while the suicide item on the HRSD includes a mix of suicidal ideation and past suicidal behavior. Second, however, the self-administered BSSI has the greatest ease of administration. Of the three measures compared in this study, the HRSD provides the most efficient method of assessing suicide along with other depression symptoms compared to the 19-item BSSI and the 5-item C-SSRS, as time and cost are often considerations when selecting assessment measures for clinical trials. Our findings support the use of the HRSD suicide item as both a suicide screening tool while the full HRSD tracks overall depression treatment outcomes that will be both time- and cost-effective if the resources for a depression clinical trial cannot support separate depression and detailed suicidal ideation scales.

Our findings provide preliminary support that suicidal ideation assessment measures, including the suicide item of the HRSD, the suicidal ideation subscale of the C-SSRS, and BSSI, provide similar estimates of suicidal ideation. Future research should not only attempt to replicate our findings but also include other suicide assessment measures such as the Collaborative Assessment and Management of Suicidality (CAMS) Suicide Status Form (Brausch et al., 2020; Jobes et al., 2004), and the suicide items on the Patient Health Questionnaire and Beck Depression Inventory (Beck et al., 1996; Kroenke et al., 2001). This research would substantially extend the present study and would lay a strong empirical groundwork for selecting a gold-standard suicide assessment for clinical research.

CONCLUSION

The current study compared the utility of three suicidal ideation assessment measures (i.e., BSSI, C-SSRS, and the HRSD suicide item) in assessing suicidal ideation in a clinical research study. Our findings show that the three measures are all equally able to detect low and very high levels of suicidal ideation. Differences in performance between the measures were modest, and each of the three items seems to be useful in assessing suicidal ideation in clinical trials, with the choice of the specific measure perhaps dictated by time and financial resources, versus the need for granularity in the interpretation of the scores.

ACKNOWLEDGEMENTS

This work was supported by NIMH MH095776, MH095780, and MH095778.

ORCID

William V. McCall https://orcid.org/0000-0001-5035-7858
Ashley R. Pate https://orcid.org/0000-0003-4917-4039
Christopher W. Drapeau https://orcid.org/0000-0002-1304-5369
Michael R. Nadorff https://orcid.org/0000-0002-8107-7514

REFERENCES

Batterham, P. J., Ftanou, M., Pirkis, J., Brewer, J. L., Mackinnon, A. J., Beautrais, A., Fairweather-Schmidt, A. K., & Christensen, H. (2015). A systematic review and evaluation of measures for suicidal ideation and behaviors in population-based research. *Psychological Assessment*, *27*(2), 501–512. https://doi.org/10.1037/pas0000053

Beck, A. T., Brown, G. K., & Steer, R. A. (1997). Psychometric characteristics of the Scale for Suicide Ideation with psychiatric outpatients. *Behavior Research and Therapy*, *35*(11), 1039–1046. https://doi.org/10.1016/S0005-7967(97)00073-9

Beck, A. T., Brown, G. K., Steer, R. A., Dahlsgaard, K. K., & Grisham, J. R. (1999). Suicide ideation at its worst point: A predictor of eventual suicide in psychiatric outpatients. *Suicide and Lifethreatening Behavior*, *29*(1), 1–9. https://doi.org/10.1111/j.1943-278X.1999.tb00758.x

Beck, A. T., Kovacs, M., & Weissman, A. (1979). Assessment of suicidal intention: The Scale for Suicide Ideation. *Journal of Consulting and Clinical Psychology*, *47*(2), 343–352. https://doi.org/10.1037/0022-006X.47.2.343

Beck, A., Steer, R., & Brown, G. (1996). *Manual for the Beck Depression Inventory-II*. Psychological Corporation.

Beck, A., Steer, R., & Ranieri, W. (1988). Scale for suicide ideation: Psychometric properties of a self-report version. *Journal of Clinical Psychiatry*, *44*(4), 499–505.

Brausch, A., O'Connor, S., Powers, J., McClay, M., Gregory, J., & Jobes, D. (2020). Validating the suicide status form for the collaborative assessment and management of suicidality in a psychiatric adolescent sample. *Suicide and Life Threatening Behavior*, *50*(1), 263–276. https://doi.org/10.1111/sltb.12587

Brown, G. (2002). *A review of suicide assessment measures for intervention with adults and older adults*. Retrieved from https://www.sprc.org/sites/default/files/migrate/library/BrownReviewAssessmentMeasuresAdultsOlderAdults.pdf

Brown, G. K., Beck, A. T., Steer, R. A., & Grisham, J. R. (2000). Risk factors for suicide in psychiatric outpatients: A 20-year prospective study. *Journal of Consulting and Clinical Psychology*, *68*(3), 371–377. https://doi.org/10.1037/0022-006X.68.3.371

Castrogiovanni, P., Matemmani, I., & Deltitio, J. (1989). Discordance of self ratings versus observer ratings in the improvement of depression: Role of locus of control and aggressive behavior. *Comprehensive Psychiatry*, *30*(3), 231–235. https://doi.org/10.1016/0010-440X(89)90043-6

Center for Drug Evaluation and Research (2012). *Guidance for industry – Suicidal ideation and behavior – Prospective assessment of occurrence in clinical trials (revision 1)*. Retrieved from https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-suicidal-ideation-and-behavior-prospective-assessment-occurrence-clinical-trials

Cochrane-Brink, K., Lofchy, J., & Sakinofsky, I. (2000). Clinical rating scales in suicide risk assessment. *General Hospital Psychiatry*, *22*, 445–451. https://doi.org/10.1016/S0163-8343(00)00106-7

Panel, D. G. (1993). *Depression in primary care (Volume2: Treatment of Major Depression; AHCPR Clinical Practice Guideline, No. 5.5)*. Retrieved from Rockville, MD.

Drapeau, C., & McIntosh, J. (2020). *U.S.A. suicide: 2018 official data*. American Association of Suicidology. Retrieved from http://www.suicidology.org

Ducher, J.-L., & Dalery, J. (2008). Correlations between Beck's suicidal ideation scale, suicide risk assessment scale RSD and Hamilton's depression rating scale. *L'encephale*, *34*(2), 132–138.

Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., Huang, X., Musacchio, K. M., Jaroszewski, A. C., Chang, B. P., & Nock, M. K. (2017). Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychological Bulletin*, *143*(2), 187–232. https://doi.org/10.1037/bul0000084

Giddens, J., Sheehan, K., & Sheehan, D. (2014). The Columbia-Suicide Severity Rating Scale (C-SSRS): Has the "gold standard" become a liability? *Innovations in Clin Neuroscience*, *11*, 66–80.

Gipson, P., Agarwala, P., Opperman, K., Horwitz, A., & King, C. (2015). Columbia-Suicide Severity Rating Scale: Predictive validity with adolescent psychiatric patients. *Pediatric Emergency Care*, *31*(2), 88–94. https://doi.org/10.1097/PEC.0000000000000225

Gutierrez, P. M., Joiner, T., Hanson, J., Avery, K., Fender, A., Harrison, T., Kerns, K., McGowan, P., Stanley, I. H., Silva, C., & Rogers, M. L. (2021). Clinical utility of suicide behavior and ideation measures: Implications for military suicide risk assessment. *Psychological Assessment*, *33*(1), 1–13. https://doi.org/10.1037/pas0000876

Gutierrez, P., Joiner, T., Hanson, J., Stanley, I., Silva, C., & Rogers, M. (2019). Psychometric properties of four commonly used suicide risk assessment measures: applicability to military treatment settings. *Military Behavioral Health*, *7*(2), 177–184. https://doi.org/10.1080/21635781.2018.1562390

Gutierrez, P., Osman, A., Barrios, F., & Kopper, B. (2001). Development and initial validation of the Self-Harm Behavior Questionnaire. *Journal of Personality Assessment*, *77*, 475–490. https://doi.org/10.1207/S15327752JPA7703_08

Hamilton, M.(1960). A rating scale for depression. *Journal of Neurol Neurosurg Psychiatry*, *23*(1), 56–62. https://doi.org/10.1136/jnnp.23.1.56. PMID: 14399272.

Hallam, K., Smith, D., & Berk, M. (2009). Differences between subjective and objective assessment of the utility of electroconvulsive therapy in patients with bipolar and unipolar depression. *Journal of Affective Disorders*, *112*, 212–218.

Hom, M., Joiner, T., & Bernert, R. (2016). Limitations of a single-item assessment of suicide attempt history: Implications for standardized suicide risk assessment. *Psychological Assessment*, *28*(8), 1026–1030. https://doi.org/10.1037/pas0000241

Jobes, D. (2006). *Managing suicidal risk: A collaborative approach*. The Guilford Press.

Jobes, D., Nelson, K., Peterson, E., Pentiuc, D., Downing, V., Francini, K., & Kiernan, A. (2004). Describing suicidality: An investigation of qualitative SSF responses. *Suicide and Life Threatening Behavior*, *34*, 99–112. https://doi.org/10.1521/suli.34.2.99.32788

Kaplan, M., Asnis, G., Sanderson, W., Keswani, L., De Lecuona, J., & Joseph, S. (1994). Suicide assessment: Clinical interview vs. self-report. *Journal of Clinical Psychiatry*, *50*(2), 294–298.

Katz, I., Barry, C., Cooper, S., Kasprow, W., & Hoof, R. (2020). Use of the Columbia-Suicide Severity Rating Scale (C-SSRS) in a large sample of Veterans receiving mental health services in the Veterans Health Administration. *Suicide and Life Threatening Behavior*, *50*(1), 111–121.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lindh, A., Dahlin, M., Beckman, K., Stromsten, L., Jokinen, J., Wiktorsson, S., & Runeson, B. (2019). A comparison of suicide risk scales in predicting repeat suicide attempt and suicide: A clinical cohort study. *Journal of Clinical Psychiatry*, *80*(6), 1–7.

McCall, W. V., Benca, R. M., Rosenquist, P. B., Youssef, N. A., McCloud, L., Newman, J. C., Case, D., Rumble, M. E., Szabo, S. T., Phillips, M., & Krystal, A. D. (2019). Reducing suicidal ideation through insomnia treatment (REST-IT): A randomized clinical trial. *American Journal of Psychiatry*, *176*, 957–965. https://doi.org/10.1176/appi.ajp.2019.19030267

Millner, A., Lee, M., & Nock, M. (2015). Single-item measurement of suicidal behaviors: Validity and consequences of misclassification. *PLoS One*, *10*(10), https://doi.org/10.1371/journal.pone.0141606

Oquendo, M. A., Halberstam, B., & Mann, J. J. (2003). Risk factors for suicidal behavior: Utility and limitations of research instruments. In M. B. First (Ed.), *Standardized evaluation in clinical practice* (pp. 103–130). APPI Press (Reprinted from: Not in File).

Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., Currier, G. W., Melvin, G. A., Greenhill, L., Shen, S. A., & Mann, J. J. (2011). The Columbia-Suicide Severity Rating Scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. *American Journal of Psychiatry*, *168*(12), 1266–1277. https://doi.org/10.1176/appi.ajp.2011.10111704

Posner, K., Oquendo, M., Gould, M., Stanley, S., & Davies, M. (2007). Columbia Classification Algorithm of Suicide Assessment (C-CASA): Classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. *American Journal of Psychiatry*, *164*, 1035–1043. https://doi.org/10.1176/ajp.2007.164.7.1035

Reynolds, W. (1991). Psychometric characteristics of the Adult Suicidal Ideation Questionnaire in college students. *Journal of Psychological Assessment*, *56*(2), 289–307. https://doi.org/10.1207/s15327752jpa5602_9

Samejima, F. (2016). Graded response models. In W. van der Linden (Ed.), *Handbook of item response theory* (Vol. *1*, pp. 123–136). CRC Press (Reprinted from: Not in File).

Sokero, T., Melartin, T., Rytsala, H., Leskela, U., Lestela-Mielonen, P., & Isometsa, E. (2003). Suicidal ideation and attempts among psychiatric patients with major depressive disorder. *Journal of Clinical Psychiatry*, *64*, 1094–1100. https://doi.org/10.4088/JCP.v64n0916

How to cite this article: McCall, W. V., Porter, B., Pate, A. R., Bolstad, C. J., Drapeau, C. W., Krystal, A. D., Benca, R. M., Rumble, M. E., & Nadorff, M. R. (2021). Examining suicide assessment measures for research use: Using item response theory to optimize psychometric assessment for research on suicidal ideation in major depressive disorder. *Suicide and Life-Threatening Behavior*, 00, 1–9. https://doi.org/10.1111/sltb.12791