Contents lists available at ScienceDirect

Journal of Psychiatric Research

journal homepage: www.elsevier.com/locate/jpsychires

Are suicide risk scales sensitive to change? A scoping review

Natalie B. Riblet [a,b,*], Sarah Matsunaga [b], Brian Shiner [a,b,c], Paula P. Schnurr [c,b], Bradley V. Watts [a,b,d]

[a] *Veterans Affairs Medical Center, Mental Health Service, 215 North Main Street, White River Junction, VT, 05009, USA*
[b] *Geisel School of Medicine at Dartmouth College, 1 Rope Ferry Road, Hanover NH, 03755, USA*
[c] *National Center for PTSD, 215 North Main Street, White River Junction, VT, 05009, USA*
[d] *Veterans Rural Health Resource Center, White River Junction VA Medical Center, White River Junction, VT, 05009, USA*

ARTICLE INFO

ABSTRACT

Introduction: Clinician- or self-administered scales are frequently used to assist in detecting risk of death by suicide and to determine the effectiveness of interventions. No recent review studies have examined whether these scales are sensitive to change. We conducted a scoping review to identify suicide risk scales that are sensitive to change.
Material and methods: We searched Medline and Excerpta Medica Database from inception through March 17, 2022, to identify randomized trials, pooled analysis, quasi-experimental studies, and cohort studies reporting on sensitivity to change of suicide risk scales. We assessed sensitivity to change by examining internal and external responsiveness. Internal responsiveness evaluates whether the scale measures changes in suicide-related symptoms in response to an intervention while external responsiveness assess whether changes in the scale correspond to changes in risk of suicide. We summarized findings across studies using descriptive analysis.
Results: Among 38 eligible scales, we identified 27 scales that included items that were modifiable to change. However, only 7 scales had been studied to determine their sensitivity to change based on internal or external responsiveness. While the results of studies suggested that 6 scales have internal responsiveness, none of the included studies confirmed that scales have external responsiveness.
Discussion: A few suicide risk scales are internally responsive and may be useful in a clinical or research setting. It is unclear, however, whether changes in scores correspond to an actual change in suicide risk. Future research should confirm the external responsiveness of scales using robust metrics including suicide mortality.

1. Background

Clinician- or self-administered scales are used in clinical and research settings to assist in assessing risk of death by suicide and to determine the effectiveness of interventions designed to decrease suicide risk (Brown; TJC, 2019). The Joint Commission requires that accredited organizations use validated scales to assess the risk of suicide among patients who are evaluated (or treated) for behavioral health conditions (TJC, 2019). For example, the Columbia Suicide Severity Rating Scale (C-SSRS) which is widely used in clinical setting to assess risk and has been translated into more than 30 languages (Salvi, 2021). The increasing use of suicide risk scales such as the C-SSRS highlights the need for research to determine whether these instruments are psychometrically sound (McGrath, 2011), meaning that the scales are both reliable and valid measures of risk of suicide in the populations where

they are applied.

A series of systematic reviews and meta-analysis have concluded that there is limited evidence establishing that clinician- or self-administered scales can improve the detection of risk of death by suicide (Carter et al., 2017; Riblet et al., 2022; Runeson et al., 2017). For example, in a recent systematic review and meta-analysis of 30 scales to identify risk of suicide in adults, Riblet et al. (2022) determined that most scales had minimal utility to detect or rule out risk of death by suicide in the settings where they are commonly applied, such as an emergency room. Riblet et al., however, noted that there were a select number of scales that may improve the detection of risk of suicide under certain circumstances. For example, the C-SSRS screen may be useful in subgroups of patients who are at high risk for suicide due to their clinical presentation but whose chief complaint does not include suicidal behavior (e. g., patients seeking treatment for alcohol use disorder) (Riblet et al.,

---

2022). A related concern that has received little attention in the literature is whether scales to detect risk of suicide are responsive or sensitive to change over time. In other words, it is unclear whether improvement (or worsening) of scores on a suicide scale correspond to a real change in risk of death by suicide, or more broadly whether the score on a scale can even change.

The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) defines sensitivity to change as 'the ability of an instrument to detect important change over time in the construct to be measured'(Mokkink et al., 2010a, pg. 742). There is a lack of agreement, however, about the best method to evaluate sensitivity to change (Husted et al., 2000). In response, Husted et al. (2000) developed a framework whereby sensitivity to change relates to the internal and external responsiveness of the scale. In addition, researchers note that the evaluation of sensitivity to change should include a concurrent examination of the test-retest reliability of the scale (Aldridge et al., 2017). If a scale has poor test-retest reliability, it will decrease the precision of the scale for measuring change (Aldridge et al., 2017).

To illustrate how Husted's framework can be applied to a psychometric scale, it is helpful to use the Posttraumatic Stress Disorder (PTSD) Checklist (PCL) as an example. The PCL is a validated scale that is commonly used to measure PTSD symptoms (Bovin et al., 2016). The PCL has good test-retest reliability (Bovin et al., 2016). Studies have found that the PCL has internal responsiveness whereby scores reliably change in response to an intervention (Marx et al., 2022). Forehand et al. (2022) have also generated evidence to suggest that the PCL has external responsiveness whereby changes in PCL scores correspond to a change in an external, clinical standard. Among 190,822 Veterans with repeated PCL scores, Forehand et al. found that patients with worsening scores over time had a roughly 25% higher rate of suicide as compared to patients with no change in scores. Moreover, patients who went into remission (final score ≤18) had a 44% lower risk of suicide versus those with scores >18.

In the case of suicide risk scales, there is limited knowledge about whether these scales are sensitive to change and the methods that researchers are using to study the sensitivity to change of these scales (Brown). Reviewing studies published through 2001, Brown observed that a few studies have generated promising results to suggest that certain suicide risk scales may be sensitive to change and may exhibit high test-retest reliability. These studies were typically intervention studies and evaluated the scale before and after an exposure to a treatment. However, newer scales such as the C-SSRS have been introduced to the field since the time that Brown's report was made available to the public.

To address current gaps in knowledge, we reviewed the available evidence on the sensitivity to change of scales to detect risk of suicide. Our review will help to inform both suicide prevention research and clinical suicide prevention practices. In the case of suicide prevention research, valid measures of change in suicide risk would avoid the need for the large sample sizes required to include death by suicide as an outcome (Sareen et al., 2014), allowing more potential interventions to be tested in a shorter amount of time. In the case of clinical suicide prevention, valid measures of change in suicide risk would aid in measurement-based care (Scott and Lewis, 2015), whereby clinicians could measure the effects of their individual interventions and determine whether they need to adjust the plan of care to decrease their patient's risk of death by suicide.

2. Methods

We conducted a scoping review. This type of literature review includes an *a priori* study protocol, a well-designed search strategy and a standardized data collection form but does not produce an overall summary estimate or formally assess the risk of bias of studies (Munn et al., 2018). We chose a scoping review approach because our review

investigates an area of research that is poorly understood (Munn et al., 2018).

Prior to initiating our review, we developed a protocol that outlined our search strategy, inclusion criteria and planned method of analysis. The protocol is posted on the protocols.io website (Integer ID: 68860; https://protocols.io/view/scoping-review-protocol-cfg4tjyw). We reported our findings according to the PRISMA Extension for Scoping Reviews (PRISMA-ScR) Guidelines (Tricco et al., 2018).

2.1. Selection of included scales

While suicide scales aim to detect risk of death by suicide, it is very challenging to assess their ability to achieve this goal using a typical study design. For example, in a recent systematic review and meta-analysis of the diagnostic accuracy of suicide scales, Riblet et al. (2022) excluded many studies because the outcome of interest did not include suicide. Nonetheless, similar to Runeson et al. (2017) and Carter et al. (2017), Riblet et al. were also able in their review to locate 38 scales that have been investigated to determine the relationship between a score on the scale and the outcome of suicide (*i.e., can a score measured at a single point in time predict whether a patient will (or will not) die by suicide*). Riblet et al. were able to include 30 of these suicide scales in a quantitative analysis of diagnostic accuracy. The remaining eight scales were excluded from that prior analysis due to insufficient data to permit quantitative analysis.

The finding that there are suicide scales that have been tested to assess their ability to predict death by suicide is intriguing. This prompts an important follow-up question, namely have these scales also been studied to ascertain whether they are responsive to change. This would have important implications for clinical practice and suicide research. Therefore, we performed a scoping review of existing studies on the sensitivity to change of the 38 suicide risk scales.

2.2. Data sources and searches

We searched Medline (via PubMed) and Excerpta Medica Database (Embase) from their inception until March 17, 2022. We made use of exploded MeSH terms and key words to create the following themes: scales to detect risk of death by suicide, suicide risk, sensitivity to change, and test-retest reliability. We relied on the Boolean term "AND" to find the intersection between the theme of scales to detect risk of death by suicide, suicide risk, and sensitivity to change as well as between the theme of scales to detect risk of death by suicide and test-retest reliability. We modified our approach as necessary to search Embase.

We found in our primary search that most of the studies reported on the Beck Scale for Suicide Ideation (BSS/SSI) and the C-SSRS. This was an expected finding because these two scales are well-known and frequently cited in the literature. Nonetheless, to minimize the possibility that our primary search strategy may have missed studies of the other, less well-known scales included in our review, we applied additional search methods. First, we searched each of the 38 scales from Riblet et al. (2022) by name in Medline and Embase. Second, we performed a reference review of included studies and prior reviews.

We applied no limitations in our search methods and considered studies for potential inclusion regardless of the language of origin.

2.3. Inclusion criteria and outcomes

We included randomized controlled trials (RCT), pooled analysis, quasi-experimental studies, and cohort studies that administered a suicide risk scale to a patient at two or more time points. We defined a suicide risk scale as an instrument that was designed (and tested) to detect risk of death by suicide (Riblet et al., 2022).

Aligned with recommendations from the literature (Husted et al., 2000), we evaluated the primary outcome, sensitivity to change, using a

multidimensional framework that considered internal and external responsiveness. In addition, we assessed test-retest reliability as a secondary outcome. Poor test-retest reliability would raise concerns about the psychometric soundness of the scale and its ability to be a valid measure of change (Aldridge et al., 2017).

We accepted any studies that reported on one or more of these outcomes of interest.

2.4. Primary outcome

We defined internal responsiveness as the ability of a scale to identify a real change in symptoms over a given time frame in response to an intervention (Husted et al., 2000). This may occur, for example, in the context of a clinical trial of a suicide prevention intervention. Internal responsiveness may be assessed using methods ranging from correlation analysis, paired t-tests or linear regression to the calculation of effect sizes such as a Guyatt's Responsiveness Index (Husted et al., 2000). We also allowed for studies that performed a shift analysis, examining whether patients shifted between categories of risk on the scale. We defined external responsiveness as the degree to which a change in a score (on a scale) relates to a change in an external clinical criterion including a measure of suicidal behavior as defined by the study (e.g., attempts, death by suicide) (Husted et al., 2000). External responsiveness may be assessed using informal methods (e.g., comparison of scores over time among patients with and without suicidal behavior) or formal methods (e.g., calculation of a receiver operator curves (ROC) to compare changes in scores over time with changes in suicide risk) (Husted et al., 2000).

Of note, we excluded from our analysis of external responsiveness the results of studies that compared a change in a score (on a scale) with a change in a score on a different scale that assessed for suicidal ideation or related symptoms (Al-Halabi et al., 2016; Ballard et al., 2015; Ducher et al., 2006; Posner et al., 2011; Price et al., 2009). We considered that these measurements addressed the convergent validity of the scale. Our approach is consistent with those of other studies (Batanero et al., 2020; Jacquemin et al., 2019).

2.5. Secondary outcome

We defined test-retest reliability, as "the systematic examination of consistency, reproducibility, and agreement among two or more measurements of the same individual, using the same [scale], under the same conditions"(Aldridge et al., 2017, pg. 208). To mitigate the risk for measurement bias, we required that reliability testing be completed within a two-week period (Polit, 2014). It is common for researchers to use a cut-off of two weeks to study test-retest reliability (Polit, 2014). The COSMIN panel also recommends that researchers choose a time interval that is most appropriate for the construct of interest (Mokkink et al., 2010b). We assumed that dynamic factors related to suicide risk (e.g., suicidal ideation) are susceptible to change. We also assumed that the likelihood of a change occurring in these factors increases with the passage of time and a real change in symptoms may occur as early as two-weeks after the prior assessment.

While test-retest reliability can be calculated in a number of ways, we preferentially reported Intraclass Correlation Coefficients (ICC) (Aldridge et al., 2017). The ICC has unique advantages over a product-moment correlation (Aldridge et al., 2017; Cicchetti, 1994). Importantly, the ICC can correct for spurious findings (Cicchetti, 1994). Using published guidelines, we considered that an ICC or correlation coefficient below 0.40 had poor reliability, a result between 0.40 and 0.59 had fair reliability, a result between 0.60 and 0.74 had good reliability and a result between 0.75 and 1.00 had excellent reliability (Cicchetti, 1994).

In addition to the ICC (or correlation coefficient), we reported the results of any paired t-test comparisons that were performed as part of test-retest reliability. In the event that an ICC (or correlation coefficient) was excellent, but the paired t-test result was also statistically significant (i.e., $p < 0.05$), we downgraded our interpretation of reliability by one level to account for the divergence in the results. We also documented whether the study had provided any potential reasons for the observed discrepancy.

2.6. Data extraction

Based on our inclusion criteria, one reviewer (NR) screened the titles and abstracts of all possibly relevant studies, excluding those that were clearly ineligible. Any studies that required full-text review to determine eligibility were then evaluated by two, independent reviewers (NR, SM) using Rayyan Software (Ouzzani et al., 2016). In the case of a disagreement between reviewers, the article was then reviewed by a third reviewer (BW) who broke the tie.

We developed a standardized data collection form for this review. Two reviewers (NR, SM) independently and in duplicate abstracted data from the included studies. Variables of interest related to study characteristics, methods of analysis, and outcomes. Any discrepancies between reviewers were resolved by involving a third reviewer (BW) who broke ties.

2.7. Data synthesis and analysis

A basic requirement for a suicide risk scale to be sensitive to change includes that the scale includes items that are modifiable or changeable. For example, a history of a suicide attempt is a non-modifiable feature that is static and cannot be changed once it has occurred. Conversely, reported suicidal ideation is a dynamic trait that is potentially modifiable and responsive to treatment. Thus, we first examined the individual items of each of the 38 scales and examined whether the scale contained any potentially modifiable items. For those scales that included modifiable items, we then described the results of studies that assessed the sensitivity to change as well as the test-retest reliability of the scale if available.

3. Results

As shown in Fig. 1, our search yielded 76 reports of studies (64 unique studies) that met our inclusion criteria. Citations of included reports are available in **Supplementary Appendix 1**). Of note, there was eight cases where a publication evaluated two scales (Ducasse et al., 2014; Ducher and Dalery, 2004; George et al., 2014; Marin et al., 2019; Morrison et al., 2016; Riblet et al., 2019; Yovell et al., 2016; Zisook et al., 2010).

Supplemental Table 1 describes the 38 scales that were included in this review. The scales range from as short as 3 items to as long as 36 items. There were 11 scales that were comprised entirely of non-modifiable items and 27 scales that included one or more items that are potentially responsive to change. Among the 27 scales that have the potential to change, however, we were only able to locate 7 scales for which researchers have generated evidence to evaluate sensitivity to change; the BSS/SSI, the C-SSRS, the InterSePT Scale for Suicidal Thinking (ISST), the Suicidal Risk Assessment Scale (RSD), the Suicide Probability Scale (SPS), the Suicide Risk Checklist (SRC), and the SUAS.

Among the 59 studies that reported on the internal or external responsiveness of one or more of these scales, we found that most were RCTs (see Table 1). The trials usually tested the scale in adult participants who were recruited from an inpatient or outpatient setting. The studies followed patients for as short as 24 h to several years. The most frequently studied scales were the BSS/SSI and C-SSRS, representing more than 12,000 patients.

3.1. Internal responsiveness

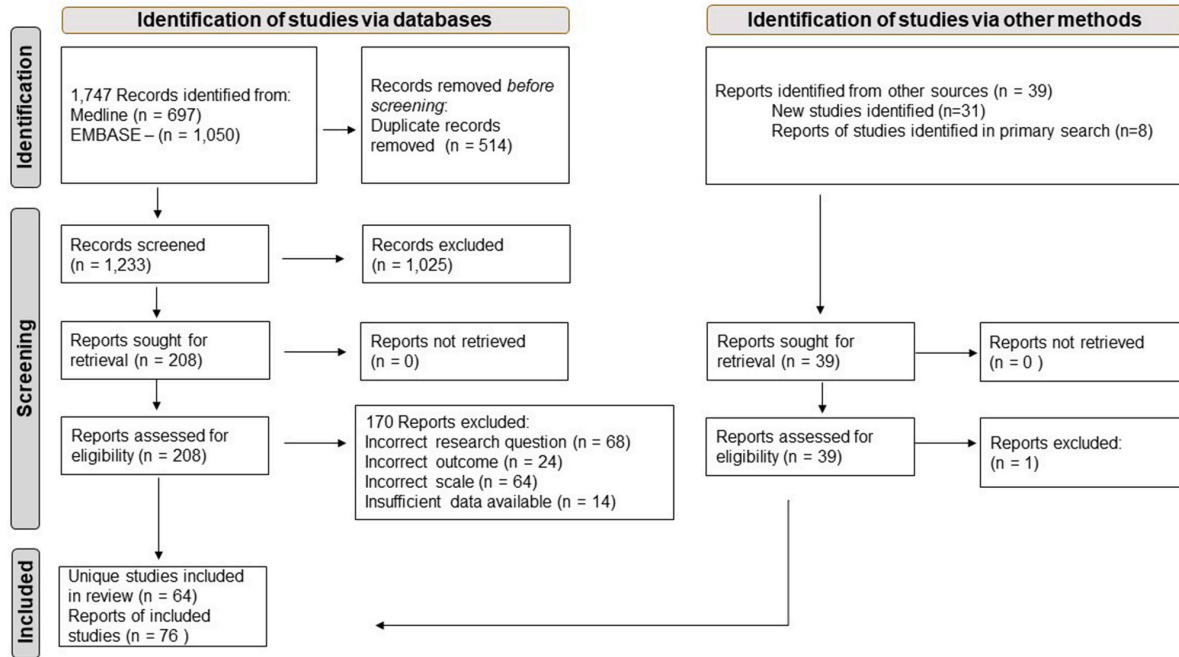As shown in Table 2, studies assessed the internal responsiveness of

Fig. 1. Prisma flow diagram.

Table 1
Characteristics of studies that have reported on internal or external responsiveness of seven suicide risk scales.[a]

Variable	Beck Scale for Suicidal Ideation/Scale for Suicidal Ideation	Columbia Suicide Severity Rating Scale	InterSePT Scale for Suicidal Thinking	The Suicidal Risk Assessment Scale, RSD	Suicide Probability Scale	Suicide Risk Checklist	Suicide Assessment Scale
N Studies	35	16	2	2	8	1	1
N Patients	2400	9665	840	211	841	53	8
Study size: Studies % (N)							
0–49 patients	51.4% (18)	50.0% (8)	0.0% (0)	0.0% (0)	37.5% (3)	0.0% (0)	100.0% (1)
50–99 patients	22.9% (8)	12.5% (2)	0.0% (0)	0.0% (0)	25.0% (2)	100.0% (1)	0.0% (0)
100–149 patients	17.1% (6)	6.3% (1)	0.0% (0)	100.0% (2)	12.5% (1)	0.0% (0)	0.0% (0)
150–199 patients	5.7% (2)	0.0% (0)	50.0% (1)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
200 + patients	2.9% (1)	31.3% (5)	50.0% (1)	0.0% (0)	25.0% (2)	0.0% (0)	0.0% (0)
Study design: Studies % (N)							
RCT	65.7% (23)	31.3% (5)	100.0% (2)	100.0% (2)	37.5% (3)	0.0% (0)	0.0% (0)
Pooled Analysis	2.9% (1)	12.5% (2)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
Quasi or open label	25.7% (9)	43.8% (7)	0.0% (0)	0.0% (0)	50.0% (4)	0.0% (0)	100.0% (1)
Cohort	5.7% (2)	12.5% (2)	0.0% (0)	0.0% (0)	12.5% (1)	100.0% (1)	0.0% (0)
Study setting: Studies % (N)							
Inpatient	37.1% (13)	18.8% (3)	0.0% (0)	0.0% (0)	12.5% (1)	100.0% (1)	0.0% (0)
Outpatient	40.0% (14)	37.5% (6)	50.0% (1)	50.0% (1)	62.5% (5)	0.0% (0)	0.0% (0)
Both	8.6% (3)	0.0% (0)	0.0% 0)	0.0% (0)	12.5% (1)	0.0% (0)	0.0% (0)
Not Reported	14.3% (5)	43.8% (7)	50.0% (1)	50.0% (1)	12.5% (1)	0.0% (0)	100.0% (1)
Age of the population: Studies % (N)							
Child &/or adolescent	0.0% (0)	25.0% (4)	0.0% (0)	0.0% (0)	37.5% (3)	0.0% (0)	0.0% (0)
Adult	88.6% (31)	75.0% (12)	100.0% (2)	100.0% (2)	50.0% (4)	100.0% (1)	0.0% (0)
Both	5.7% (2)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
Not Reported	5.7% (2)	0.0% (0)	0.0% (0)	0.0% (0)	12.5% (1)	0.0% (0)	100.0% (1)
Maximum length of follow-up: Studies % (N)							
≤7 days	17.1% (6)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (1)
8–91 days	31.4% (11)	6.3% (1)	50.0% (1)	50.0% (1)	50.0% (4)	0.0% (0)	0.0% (0)
92–182 days	17.1% (6)	50.0% (8)	0.0% (0)	0.0% (0)	12.5% (1)	0.0% (0)	0.0% (0)
183–365 days	22.9% (8)	31.3% (5)	0.0% (0)	0.0% (0)	25.0% (2)	0.0% (0)	0.0% (0)
>365 days	8.6% (3)	12.5% (2)	50.0% (1)	50.0% (1)	12.5% (1)	0.0% (0)	0.0% (0)
Not Reported	2.9% (1)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	100.0% (1)	100.0% (1)

[a] There were eight studies that examined two scales and thus, are referenced several times in the table.

Table 2
Internal responsiveness of suicide risk scales.[a]

	Do scores change post exposure to an intervention aimed at reducing suicide risk?				Do scores change post exposure to any intervention (excluding suicide prevention strategies)?				
Suicide Risk Scale	BSS/SSI	C-SSRS	SPS	SUAS	BSS/SSI	C-SSRS	ISST	RSD	SPS
Total Studies	26	7	5	1	9	8	1	2	3
Study performed an unadjusted (e.g., paired t-test) or adjusted analysis (e.g., linear regression): Studies % (N)									
N Studies	26	7	5	1	8	2	1	2	2
No change	23.1 (6)	28.6 (2)	40.0 (2)	100.0 (1)	62.5 (5)	50.0 (1)	0.0 (0)	0.0 (0)	100.0 (2)
Sig. change	76.9 (20)	71.4 (5)	60.0 (3)	0.0 (0)	37.5 (3)	50.0 (1)	100.0 (1)	100.0 (2)	0.0 (0)
Study performed a shift analysis: Studies % (N)									
N Studies	4	1	1	0	0	5	1	0	0
<10% shift in category	0.0 (0)	0.0 (0)	0.0 (0)	–	–	0.0 (0)	0.0 (0)	–	–
10–49% shift in category	25.0 (1)	0.0 (0)	100.0 (1)	–	–	80.0 (4)	0.0 (0)	–	–
50–74% shift in category	50.0 (2)	100.0 (1)	0.0 (0)	–	–	0.0 (0)	100.0 (1)	–	–
>75% shift in category	25.0 (1)	0.0 (0)	0.0 (0)	–	–	20.0 (1)	0.0 (0)	–	–
Study calculated effect sizes (Cohen's d; Hedges' g): Studies % (N)									
N Studies	10	1	0	0	0	0	0	0	0
Small (0.20–0.49)	30.0 (3)	0.0 (0)	–	–	–	–	–	–	–
Medium (0.50–0.79)	40.0 (4)	0.0 (0)	–	–	–	–	–	–	–
Large (≥0.80)	30.0 (3)	100.0 (1)	–	–	–	–	–	–	–
Study calculated a reliable reduction in scorers using the Jacobson and Truax Reliable Change Index[b]: Studies % (N)									
N Studies	2	0	0	0	0	0	0	0	0
<10% reliable reduction	0.0 (0)	–	–	–	–	–	–	–	–
10–49% reliable reduction	50.0 (1)	–	–	–	–	–	–	–	–
50–74% reliable reduction	50.0 (1)	–	–	–	–	–	–	–	–
>75% reliable reduction	0.0 (0)	–	–	–	–	–	–	–	–
Study measured correlation between scores before and after exposure to intervention: Studies % (N)									
N Studies	0	0	0	0	1	0	0	0	0
None (<0.20)	–	–	–	–	0.0 (0)	–	–	–	–
Weak (0.20–0.39)	–	–	–	–	0.0 (0)	–	–	–	–
Moderate (0.40–0.59)	–	–	–	–	100.0 (1)	–	–	–	–
Strong (≥0.6)	–	–	–	–	0.0 (0)	–	–	–	–
Study measured correlation between exposure to intervention and standardized change score: Studies % (N)									
N Studies	0	0	0	0	0	0	0	0	1
No correlation	–	–	–	–	–	–	–	–	0.0 (0)
Sig. correlation	–	–	–	–	–	–	–	–	100.0 (1)
Study measured change in scores but did not describe analytic methods: Studies % (N)									
N Studies	0	0	0	0	0	1	0	0	0
No change	–	–	–	–	–	100.0 (1)	–	–	–
Evidence of change	–	–	–	–	–	0.0 (0)	–	–	–

BSS/SSI = Beck Scale for Suicidal Ideation/Scale for Suicidal Ideation; C-SSRS = Columbia Suicide Severity Rating Scale; ISST = InterSePT Scale for Suicidal Thinking; N = number; % = proportion; RSD = The Suicidal Risk Assessment Scale, RSD; SPS = Suicide Probability Scale; SRC = Suicide Risk Checklist; SUAS = Suicide Assessment Scale.

[a] There were eight studies that examined two scales and thus, are referenced several times in the table.

[b] Studies defined a reliable reduction in scores as either a score of >5.84 points or >6.48 points.

six scales (BSS/SSI, C-SSRS, ISST, RSD, SPS, and SUAS) in response to a suicide prevention strategy (N = 39) or other intervention (N = 23). Among the six scales, there was rather robust evidence that the BSS/SSI and C-SSRS exhibit internal responsiveness. For example, 25 studies observed that BSS/SSI or C-SSRS scores changed significantly in response to a suicide prevention. Eight studies were able to show that the size of the change in the BSS/SSI or C-SSRS was medium or large. Furthermore, two studies (Roush et al., 2021; Van Spijker et al., 2012) reported that a sizable proportion of patients experienced a reliable change in BSS scores based on the Jacobson and Truax's reliable change index (Jacobson and Truax, 1991).

3.2. External responsiveness

We located four studies that examined the external responsiveness of four suicide risk scales (see Table 3). The studies examined scales in various populations ranging from patients with a history of a suicide attempt (N = 1) to psychiatrically hospitalized patients (N = 2) to patients with schizophrenia spectrum disorder (N = 1). In general, studies suggested that the C-SSRS, the SPS, and the SRC may be sensitive to change, although none of the findings were conclusive. The results of a single study of the ISST also yielded inconclusive evidence. While scores significantly increased over time in those with suicidal behavior, the scale had low diagnostic accuracy (e.g., sensitivity between 9 and 15%). The authors found no clear differences in scores in those who died by

suicide, but the number of deaths was very low, precluding a robust analysis.

We did not locate any studies that examined the external responsiveness of the BSS/SSI, RSD, or SUAS.

3.2.1. Test-retest reliability

We identified six studies that reported on test-retest reliability for five of the seven scales (i.e., BSS/SSI, C-SSRS, ISST, SPS, and SUAS) (see Table 4). The test-retest reliability of the C-SSRS, ISST, SPS, and SUAS was generally excellent. The test-retest reliability of the BSS on the other hand ranged from fair to good. Beck and Steer (1991) hypothesized that patients' scores on the BSS had improved because they had received inpatient treatment. Pinninti et al. (2002) highlighted two concerns. First, seven of the 15 subjects had scored a zero on the baseline and follow-up assessments. Thus, the correlations may have been falsely high. Second, the patients' symptoms have may have improved because of inpatient treatment.

Of note, Fosse et al. (2017) described that they had performed reliability testing on the SRC. Among the 53 patients who had two or more assessments (and completed the SRC), Fosse et al. found that the correlation between total scores was moderate (i.e., r = 0.40, p = 0.003). The timing of the measurement, however, was unclear. The assessments were completed at the time of hospital admission but the amount of time that elapsed between admissions was not stated.

We did not find any studies that reported on the test-retest reliability

Table 3
External responsiveness of suicide risk scales.[a]

Author/ Year	N	Population	External Outcome	Key Findings	Conclusions
Columbia Suicide Severity Rating Scale (C-SSRS)					
Posner et al., 2011	124	Adolescents with history of suicide attempt	Columbia Suicide History Form (CSHF); Independent evaluation board	Using past week and since last visit assessments, there was full agreement with the CSHF for interrupted and actual attempts and moderate agreement for aborted attempts (kappa = 0.66, 95%CI: 0.23–1.00). There was high agreement with results of the evaluation board for attempt and interrupted attempts (kappa = 0.88, 95%CI: 0.77–0.98). The authors noted that one suicide case was excluded from the analysis because it was rated by the board after the study ended.	C-SSRS behavior subscale may be sensitive to change
InterSePT Scale for Suicidal Thinking (ISST)					
Ayer 2008	642	Adults with Schizophrenia or Schizo-affective Disorder	Suicide attempt or hospitalized to prevent suicidal behavior	Compared to those without an event, patients with an event experienced significant increases in scores between assessments (p0.02). The diagnostic accuracy of the ISST, however, was poor. While an increase of ≥6 points on the ISST yielded nearly 100% specificity for predicting an event, the sensitivity at this cutoff was low (i.e., 9–15%). The positive predictive values ranged between 83 and 100%.	Sensitivity to change of ISST is inconclusive
Ayer 2008	642	Adults with Schizophrenia or Schizo-affective disorder	Death by suicide	There were no clear trends over time in the ISST scores for the eight patients who died by suicide. One patient had no improvement in scores over 16-months. The remaining seven patients had modest reductions in scores after first assessment and/or denied SI at baseline and had no worsening in scores at later assessment	Sensitivity of change of ISST is inconclusive
Suicide Probability Scale (SPS)					
Eltz 2007	226	Adolescent psychiatric inpatients	Number of readmissions for repeat suicidal behavior	Negative association between SPS scores during a patient's hospitalization and readmission, r = −0.21, p = 0.05. In other words, patients with improvement in scores were less likely to be readmitted, while patients with worsening scores were more likely to be readmitted.	SPS may be sensitive to change
Suicide Risk Checklist (SRC)					
Fosse et al., 2017	53	Adult psychiatric inpatients	Death by suicide	Total scores were higher among suicides versus controls at both assessments (p < 0.04). In an analysis of the subscales (i.e., part A and B), only part B was higher among suicides versus controls, and this was only true for the last assessment (p = 0.008). A trend analysis suggested that part A was higher among suicides versus controls at last assessment, p = 0.098. There was also a trend to higher scores on Part A and B among suicides versus controls at first assessment, p = 0.088 and 0.08, respectively.	SRC may be sensitive to change

[a] No studies reported on the external responsiveness of the Beck Scale for Suicidal Ideation, the Suicide Assessment Scale, or The Suicide Risk Assessment Scale, RSD.

of the RSD.

4. Discussion

Suicide is a major public health problem and there is high motivation to prevent suicide in the population. Nonetheless, suicide is an uncommon event, and it is very difficult for clinicians and researchers to reliably predict who will go on to die by suicide. In a clinical setting, suicide scales are routinely used by providers to assess and to document near- or long-term risk of suicide. If these measures, however, do not accurately assess risk (or changes in risk over time), this poses a serious problem for the healthcare system. Similarly, because researchers commonly use suicide scales to draw conclusions about the efficacy of interventions, it is of high concern if these scales do not accurately reflect the effect of the intervention on the risk of suicide.

We conducted a scoping review of clinician or self-administered suicide risk scales. Our goal was to identify and describe the results of studies that have assessed the ability of suicide risk scales to be sensitive to change. We initially located 38 scales that were eligible to be included in our review. We found that eleven of these scales had no items that could change over time. The remaining 27 scales varied widely in the number of included items that were modifiable, ranging from 100% of items to 12.5% of items. We determined that only 7 suicide risk scales have been tested to determine their sensitivity to change. There was evidence that these scales are sensitive to change and may be applied in a clinical or research setting. This conclusion, however, is based primarily on studies that assessed the internal responsiveness of scales. The

literature regarding the external responsiveness of scales was quite limited. While changes in the scale may be correlated with changes in other measures of suicide risk, the evidence to support an association between changes in scores and death by suicide was lacking.

Despite the overwhelming and rightful focus on suicide as a public health emergency, there has been limited focus on the basic tools needed to assist clinicians and researchers in measuring suicide risk over time. The seminal work in this area by Brown was actually posted on a website. While our scoping review diverged somewhat from Brown's methods and we reported on a different set of scales, our basic conclusions are closely aligned: there are several suicide risk scales that seem to have utility in tracking suicide risk over time and more study is needed to better demonstrate the relationship between change in scores and risk of suicide mortality.

Our finding that a large majority of studies determined that the BSS/SSI and C-SSRS exhibit internal responsiveness is somewhat reassuring. Our conclusions, however, are based primarily on the results of intervention studies that compared mean changes over time using methods such as paired-t-tests, analysis of variance (ANOVA), and Cohen's *d*. Husted et al. (2000) emphasize that these methods provide limited insight into the sensitivity to change of scales. In fact, these methods are imperfect indicator of whether a change has occurred at the level of the patient (Husted et al., 2000). We were only able to locate two studies of the BSS that used more robust metrics of internal responsiveness at the individual level (i.e., the Jacobson and Truax's reliable change index) (Jacobson and Truax, 1991).

Related to these concerns, we described the internal responsiveness

Table 4
Test-retest reliability of suicide risk scales.[a]

Author/Year	N	Timing of assessment	Calculation of the Intra Class Correlation (ICC) or Pearson's Correlation Coefficient (r)	Results of paired *t*-test	Interpretation of findings
Beck Scale for Suicide Ideation (BSS)					
Beck 1991	60	1 week later	The study enrolled 339 psychiatric inpatients. In a subsample of 60 patients, the correlation between BSS at baseline and at follow-up was r = 0.54.	T(57) = 4.25, p < 0.001	Fair
Pinninti et al., 2002	15	1 week later	The study enrolled 130 psychiatric inpatients. In a subsample of 15 patients, the correlation between BSS at baseline and at follow-up was r = 0.88.	T(14) = 2.75, p < 0.05	Good
Columbia Suicide Severity Rating Scale (C-SSRS)					
Gwaltney et al., 2017	86	Same day; three assessments 20–25 min apart; separated by lexicon task	In a sample of adult psychiatric inpatients and hospital employees, the ICC between the electronic C-SSRS (lifetime most severe suicidal ideation) at baseline and final assessment was 0.87 using an Interactive Voice Response version and 0.87 using a text-based version. In a sample of 67 patients who endorsed lifetime suicidal ideation, the ICC between the electronic C-SSRS (past six months most severe suicidal ideation) at baseline and at final assessment was 0.72 using an Interactive Voice Response version and 0.84 using a text-based version.	Not reported	Good to excellent
InterSePT Scale for Suicidal Thinking (ISST)					
Hammoudeh 2016	22	3 days later	The study enrolled 199 adults including 100 adults with schizophrenia and 99 controls. In a subsample of 22 patients with schizophrenia, the ICC between scores on the Arabic version of the ISST at baseline and at follow-up was 0.93.	Not reported	Excellent
Suicide Probability Scale (SPS)					
Cull 1982 (cited in Range, 1989)	1158[b]	10 days later	The study enrolled 1158 patients including 562 adolescents and adults with no prior psychiatric history, 260 psychiatric inpatients and 336 outpatients with a recent history of a suicide attempt. The authors found that the correlation between SPS at baseline and at follow-up was r = 0.94. *Note:* An alternative form of the SPS was tested in 47 college students and found to have a r = 0.90 at a time interval of 30 min.	Not reported	Excellent
Suicide Assessment Scale (SUAS)					
Koldsland 2012	52	1 week later	In a sample of 52 adult outpatients with major depression, bipolar disorder and/or personality disorder, the ICC between SUAS at baseline and at follow-up was 0.95.	Not reported	Excellent

[a] No studies reported on the ≤2 week test-retest reliability of the Suicide Risk Checklist or The Suicide Risk Assessment Scale, RSD.
[b] The publication did not specify the specific number of patients that contributed to the test-retest reliability analysis.

of scales using change statistics. There are likely better ways, however, to measure internal responsiveness. We found that the studies that reported on change statistics varied in size from small (<20) to large (>100) and used parametric or non-parametric tests. In some cases, the studies were looking for differences in pre-post scores in a single intervention arm, while in other cases they were looking for differences in scores between arms (e.g., intervention versus treatment as usual). A lack of a statistically significant finding could reflect problems with the scale, the intervention, and/or the sample itself. The usefulness of a statistical test depends highly on whether the test is applied correctly in the context of the study. Because our scoping review was not designed to examine these factors, we cannot rule out that these issues may have played a role in our findings.

A suicide scale could exhibit internal responsiveness and yet, it's possible that an improvement (or worsening) in the scores over time may not actually change the individual's risk of suicide. To address this concern, we examined whether a change in a score on a suicide scale related to an external clinical criterion such as death by suicide, We identified four studies that assessed the external responsiveness of four scales including the CSSR-S, ISST, SPS, and SRC. A few of these studies yielded promising trends in favor of the external responsiveness the C-SSRS, SPS and SRC. None of the studies, however, were able to make a definitive determination about whether the scales exhibit external responsiveness. It is critical that future research examine this question.

Reassuringly, we found that the test-retest reliability of studied scales was generally excellent. The test-retest reliability of the BSS, however, is more limited. A unique challenge in assessing the test-retest reliability of suicide risk scales includes that these scales are measuring symptoms that are dynamic and may be responsive to an intervention. As such, symptoms may change more frequently than is accounted for in a typical testing window. For example, Beck and Steer (1991) pointed

out that patients were exposed to hospital treatment during the one-week testing period. As a result, patients may have experienced a significant improvement in symptoms and the strength of the correlation between the baseline and follow-up assessment may have been diminished (Beck and Steer, 1991). Other researchers have also pointed out that suicidal ideation may vary considerably even during the course of a single day (Witte et al., 2005). To address these concerns, Gwaltney et al. (2017) administered the C-SSRS scale three-times on the same day and used a lexical decision task with refreshment break to separate the testing administrations. These authors found that the C-SSRS had good to excellent test-retest reliability.

A strength of our work includes that we are the first study to comprehensively review the literature to identify (and describe) studies examining the sensitivity to change of suicide risk scales. We used a comprehensive framework to assess sensitivity to change (Aldridge et al., 2017; Husted et al., 2000). Our review also highlights the possibility that improved metrics could be developed potentially by combining existing, but non-overlapping items on the available suicide risk scales. This is feasible especially in light of the relative brevity of all the scales.

Our work, however, is not without limitations both in our approach and the available data to support our approach. Although we conducted a thorough search of the literature, it is possible that we overlooked other relevant studies. In our review, we observed that included studies tested scales in populations with varying degrees of symptoms and a range of psychopathologies. In psychiatric populations, in particular, it is important to consider the amount of within-subject variation and to account for this variation when assessing the performance of scales. For example, in a study of 30 patients with schizophrenia or schizoaffective disorder, Lucas and Wade (2001) used an inflated alpha level of 0.07 to address issues related to within-subject variation.

We chose to focus our review exclusively on suicide risk scales that have been previously studied to determine their ability to detect risk of death by suicide based on a single assessment. As a result, we excluded other suicide scales that have been tested to determine their ability to detect other measures of suicidal behavior (e.g., suicide attempts). Studies have found that some scales such as the Computerized Adaptive Test Suicide Scale (CAT-SS) are internally responsive and can predict non-fatal, suicide-related outcomes as measured by the C-SSRS (i.e., active suicidal ideation with plan and intent, suicide attempts, interrupted attempts, aborted attempts, self-interrupted attempts, or preparatory acts or behaviors) (Brenner et al., 2022; Grunebaum et al., 2021). These findings highlight the need for a future review to examine how changes in scores in these other suicide scales relate to external measures of suicidal behavior.

We were only able to identify studies of the sensitivity to change of seven of the 27 eligible scales. Most of the included studies reported on the BSS/SSI or C-SSRS. This suggests that in general there has been very limited reporting of the sensitivity to change of suicide scales. A related concern includes the fact that very few studies assessed the external responsiveness of suicide risk scales. This highlights a particular challenge in suicide research whereby the reference standard, death by suicide, is rare. Thus, studies require very large sample sizes or need to follow patients for a long period of time (Sareen et al., 2014). The costs of carrying out such a study may necessarily be prohibitive for may research teams.

5. Conclusion

In summary, our scoping review illuminates there are a few suicide risk scales, such as the BSS/SSI and C-SSRS that are sensitive to change based on internal responsiveness. These scales may be useful in a clinical or research setting. It remains unclear, however, whether changes in scores on suicide risk scales correspond to an actual change in suicide risk. Additional work validating suicide risk scales is warranted including the challenging process of validation to suicide mortality.

Financial disclosure

Grant funding

Prior presentations

None.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpsychires.2023.03.007.

References

Aldridge, V.K., Dovey, T.M., Wade, A., 2017. Assessing test-retest reliability of psychological measures: persistent methodological problems. Eur. Psychol. 22 (4), 207–218. https://doi.org/10.1027/1016-9040/a000298.

Al-Halabí, S., Sáiz, P.A., Burón, P., Garrido, M., Benabarre, A., Jiménez, E., Cervilla, J., Navarrete, M.I., Díaz-Mesa, E.M., García-Álvarez, L., Muñiz, J., Posner, K., Oquendo, M.A., García-Portilla, M.P., Bobes, J., 2016. Validation of a Spanish version of the columbia-suicide severity rating scale (C-SSRS). Rev. Psiquiatía Salud Ment. 9 (3), 134–142.

Ballard, E.D., Luckenbaugh, D.A., Richards, E.M., Walls, T.L., Brutsche, N.E., Ameli, R., Niciu, M.J., Vande Voort, J.L., Zarate Jr., C.A., 2015. Assessing measures of suicidal ideation in clinical trials with a rapid-acting antidepressant. J. Psychiatr. Res. 68, 68–73. https://doi.org/10.1016/j.jpsychires.2015.06.003.

Batanero, C.D., Sanchez-Garcia, M., Fernandez-Calderon, F., Lozano, O.M., 2020. Decision making in substance use disorder patients: reliability, convergent validity and sensitivity to change of a new behavioral measure. J. Clin. Exp. Neuropsychol. 42 (10), 998–1009.

Beck, A.T., Steer, R.A., 1991. BSI, Beck Scale for Suicide Ideation: Manual. Psychological Corp, New York.

Bovin, M.J., Marx, B.P., Weathers, F.W., Gallagher, M.W., Rodriguez, P., Schnurr, P.P., Keane, T.M., 2016. Psychometric properties of the PTSD checklist for diagnostic and statistical manual of mental disorders – fifth edition (PCL-5) in veterans. Psychol. Assess. 28 (11), 1379–1391.

Brenner, L.A., Betthauser, L.M., Penzenik, M., Bahraini, N., Gibbons, R.D., 2022. Validation of a computerized adaptive test suicide scale (CAT-SS) among United States military veterans. PLoS One 17 (1), e0261920.

Brown, G.K. A review of suicide assessment measures for intervention research with adults and older adults. Technical report submitted to the National Institutes of Mental Health. https://www.sprc.org/resources-programs/review-suicide-assessment-measures-intervention-research-adults-older-adults. (Accessed 9 December 2021). accessed.

Carter, G., Milner, A., McGill, K., Pirkis, J., Kapur, N., Spittal, M.J., 2017. Predicting suicidal behaviours using clinical instruments: systematic review and meta-analysis of positive predictive values for risk scales. Br. J. Psychiatry 210 (6), 387–395. https://doi.org/10.1192/bjp.bp.116.182717.

Cicchetti, D.V., 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol. Assess. 6 (4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284.

Ducasse, D., Rene, E., Beziat, S., Guillaume, S., Courtet, P., Olie, E., 2014. Acceptance and commitment therapy for management of suicidal patients: a pilot study. Psychother. Psychosom. 83, 374–376. https://doi.org/10.1159/000365974.

Ducher, J.L., Dalery, J., 2004. Concurrent validation of the suicidal risk assessment scale (R.S.D.) with the Beck's suicidal ideation scale. [French]. [Validation concourante de l'echelle d'evaluation du risque suicidaire (R.S.D.) avec l'echelle d'ideation suicidaire de Beck.]. Encephale 30 (3), 249–254. https://doi.org/10.1016/s0013-7006%2804%2995436-x.

Ducher, J.L., Terra, J.L., 2006. Is the suicidal risk assessment scale RSD of predictive value? Encephale 32 (5), 738–745. https://doi.org/10.1016/s0013-7006(06)76226-1.

Forehand, J.A., Dufort, V., Gradus, J.L., Maguen, S., Watts, B.V., Jiang, T., Holder, N., Shiner, B., 2022. Association between post-traumatic stress disorder severity and death by suicide in US military veterans: retrospective cohort study. BJPsych 221, 676–682.

Fosse, R., Ryberg, W., Carlsson, M.K., Hammer, J., 2017. Predictors of suicide in the patient population admitted to a locked-door psychiatric acute ward. PLoS One 12 (3), e0173958.

George, M.S., Raman, R., Benedek, D.M., Pelic, C.G., Grammer, G.G., Stokes, K.T., Schmidt, M., Spiegel, C., DeAlmeida, N., Beaver, K.L., Borckardt, J.J., Sun, X., Jain, S., Stein, M.B., 2014. A two-site pilot randomized 3 day trial of high dose left prefrontal repetitive transcranial magnetic stimulation (rTMS) for suicidal inpatients. Brain Stimul. 7, 421–431.

Grunebaum, M.F., Mann, J.J., Galfalvy, H.C., Gibbons, R.D., 2021. Computerized-Adaptive vs. traditional ratings of depression and suicidal thoughts: an assay sensitivity pilot study in a ketamine clinical trial. Front. Psychiatr. 12, 602976.

Gwaltney, C., Mundt, J.C., Greist, J.H., Paty, J., Tiplady, B., 2017. Interactive voice response and text-based self-report versions of the electronic columbia-suicide severity rating scale are equivalent. Innov. Clin. Neurosci. 14, 17–23.

Husted, J.A., Cook, R.J., Farewell, V.T., Gladman, D.D., 2000. Methods for assessing responsiveness: a critical review and recommendations. J. Clin. Epidemiol. 53, 459–468.

Jacobson, N.S., Truax, P., 1991. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J. Consult. Clin. Psychol. 59 (1), 12–19. https://doi.org/10.1037//0022-006x.59.1.12.

Jacquemin, L., Mertens, G., Van de Heyning, P., Vanderveken, O.M., Topsakal, V., De Hertogh, W., Michiels, S., Van Rompaey, V., Gilles, A., 2019. Sensitivity to change and convergent validity of the tinnitus functional index (TFI) and the tinnitus questionnaire (TQ): clinical and research perspectives. Hear. Res. 382, 1–8.

Lucas, S., Wade, T., 2001. An examination of the power of the voices in predicting the mental state of people experiencing psychosis. Behav. Change 18 (1), 51–57.

Marin, L., Sullivan, S., Spears, A.P., Goodman, M., 2019. Project Life Force-Geriatric": a novel suicide safety planning group treatment. Am. J. Geriatr. Psychiatr. 27 (3 Supplement), S172–S173. https://doi.org/10.1016/j.jagp.2019.01.127.

Marx, B.P., Lee, D.J., Norman, S.B., Bovin, M.J., Sloan, D.M., Weathers, F.W., Keane, T. M., Schnurr, P.P., 2022. Reliable and clinical significant change in the Clinician-Administered PTSD Scale for DSM-5 and PTSD Checklist for DSM-5 among male veterans. Psychol. Assess. 34 (2), 197–203.

McGrath, M.C., 2011. Psychometrics. In: Goldstein, S., Naglieri, J.A. (Eds.), Encyclopedia of Child Behavior and Development. Springer, Boston, MA, pp. 1190–1191. https://doi.org/10.1007/978-0-387-79061-9_2306.

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M., de Vet, H.C., 2010a. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J. Clin. Epidemiol. 63 (7), 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006.

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M., de Vet, H.C.W., 2010b. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual. Life Res. 19 (4), 539–549. https://doi.org/10.1007/s11136-010-9606-8.

Morrison, A.P., Law, H., Barrowclough, C., Bentall, R.P., Haddock, G., Jones, S.H., Kilbride, M., Pitt, E., Shryane, N., Tarrier, N., Welford, M., Dunn, G., 2016. Psychological Approaches to Understanding and Promoting Recovery in Psychosis and Bipolar Disorder: a Mixed-Methods Approach. NIHR Journals Library, Southhampton (UK). https://doi.org/10.3310/pgfar04050.

Munn, Z., Peters, M.D.J., Stern, C., Tufanaru, C., McArthur, A., Aromataris, E., 2018. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med. Res. Methodol. 18 (1), 143. https://doi.org/10.1186/s12874-018-0611-x.

Ouzzani, M., Hammady, H., Fedorowicz, Z., Elmagarmid, A., 2016. Rayyan - a web and mobile app for systematic reviews. Syst. Rev. 5, 210. https://doi.org/10.1186/s13643-016-0384-4.2016.

Pinnini, N., Steer, R.A., Rissmiller, D.J., Nelson, S., Beck, A.T., 2002. Use of the beck scale for suicide ideation with psychiatric inpatients diagnosed with schizophrenia, schizoaffective, or bipolar disorders. Behav. Res. Ther. 40, 1071–1079.

Polit, D.F., 2014. Getting serious about test–retest reliability: a critique of retest research and some recommendations. Qual. Life Res. 23 (6), 1713–1720. https://doi.org/10.1007/s11136-014-0632-9.

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., Mann, J.J., 2011. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am. J. Psychiatr. 168 (12), 1266–1277. https://doi.org/10.1176/appi.ajp.2011.10111704.

Price, R.B., Nock, M.K., Charney, D.S., Mathew, S.J., 2009. Effects of intravenous ketamine on explicit and implicit measures of suicidality in treatment-resistant depression. Biol. Psychiatr. 66 (5), 522–526.

Riblet, N., Matsunaga, S., Lee, Y., Young-Xu, Y., Shiner, B., Schnurr, P.P., Levis, M., Watts, B.V., 2022. Tools to detect risk of death by suicide: a systematic review and meta-analysis. J. Clin. Psychiatry 84 (1), 21r14385.

Riblet, N.B., Shiner, B., Schnurr, P., Bruce, M.L., Wasserman, D., Cornelius, S., Scott, R., Watts, B.V., 2019. A pilot study of an intervention to prevent suicide after psychiatric hospitalization. J. Nerv. Ment. Dis. 207 (12), 1031–1038. https://doi.org/10.1097/nmd.0000000000001061.

Roush, J.F., O'Brien, K.M., Ruha, A.L., 2021. Evaluating a recovery-oriented intensive outpatient program for Veterans at risk for suicide. Crisis 42 (3), 218–224. https://doi.org/10.1027/0227-5910/a000703.

Runeson, B., Odeberg, J., Pettersson, A., Edbom, T., Jildevik Adamsson, I., Waern, M., 2017. Instruments for the assessment of suicide risk: a systematic review evaluating the certainty of the evidence. PLoS One 12 (7), e0180292. https://doi.org/10.1371/journal.pone.0180292.

Salvi, J., 2021. Columbia-suicide severity rating scale (C-SSRS). Emerg. Med. Pract. 21 (5), CD3–CD4.

Sareen, J., Isaak, C., Katz, L.Y., Bolton, J., Enns, M.W., Stein, M.B., 2014. Promising strategies for advancement in knowledge of suicide risk factors and prevention. Am. J. Prev. Med. 47, S257–S263.

Scott, K., Lewis, C.C., 2015. Using measurement-based care to enhance any treatment. Cognit. Behav. Pract. 22 (1), 49–59. https://doi.org/10.1016/j.cbpra.2014.01.010.

The Joint Commission (TJC), 2019. National patient safety goal for suicide prevention. R3 Report. https://www.jointcommission.org/-/media/tjc/documents/standards/r3-reports/r3_18_suicide_prevention_hap_bhc_cah_11_4_19_final1.pdf. (Accessed 9 December 2021). accessed.

Tricco, A.C., Lillie, E., Zarin, W., O'Brien, K.K., Colquhoun, H., Levac, D., Moher, D., Peters, M.D.J., Horsley, T., Weeks, L., Hempel, S., Akl, E.A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M.G., Garritty, C., Lewin, S., Godfrey, C.M., Macdonald, M.T., Langlois, E.V., Soares-Weiser, K., Moriarty, J., Clifford, T., Tunçalp, O., Straus, S.E., 2018. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann. Intern. Med. 169 (7), 467–473. https://doi.org/10.7326/m18-0850.

Van Spijker, B.A.J., Majo, M.C., Smit, F., van Straten, A., Kerkhof, A.J.F.M., 2012. Reducing suicidal ideation: cost-effectiveness analysis of a randomized controlled trial of unguided web-based self-help. J. Med. Internet Res. 14 (5), e141. https://doi.org/10.2196/jmir.1966.

Witte, T.K., Fitzpatrick, K.K., Joiner Jr., T.E., Schmidt, N.B., 2005. Variability in suicidal ideation: a better predictor of suicide attempts than intensity or duration of ideation? J. Affect. Disord. 88 (2), 131–136. https://doi.org/10.1016/j.jad.2005.05.019.

Yovell, Y., Bar, G., Mashiah, M., Baruch, Y., Briskman, I., Asherov, J., Chang, A., Rigbi, A., Panksepp, J., 2016. Ultra-low-dose buprenorphine as a time-limited treatment for severe suicidal ideation: a randomized controlled trial. Am. J. Psychiatr. 173 (5), 491–498.

Zisook, S., Kasckow, J.W., Lanouette, N.M., Golshan, S., Fellows, I., Vahia, I., Mohamed, S., Rao, S., 2010. Augmentation with citalopram for suicidal ideation in middle-aged and older outpatients with schizophrenia and schizoaffective disorder who have subthreshold depressive symptoms: a randomized controlled trial. J. Clin. Psychiatry 71 (7), 915–922.