Published in final edited form as:
J Clin Psychiatry. ; 84(1): . doi:10.4088/JCP.21r14385.

Tools to Detect Risk of Death by Suicide: A Systematic Review and Meta-analysis

Natalie B. Riblet, MD MPH[1,2], **Sarah Matsunaga, BA**[3], **Younji Lee, BA**[3], **Yinong Young-Xu, ScD**[1,4], **Brian Shiner, MD MPH**[1,2,5], **Paula P. Schnurr, PhD**[5,4], **Maxwell Levis, PhD**[1,4], **Bradley V. Watts, MD MPH**[1,4,6]

[1]Veterans Affairs Medical Center, Mental Health Service, 215 North Main Street, White River Junction, VT 05009 USA

[2]Geisel School of Medicine at Dartmouth College, Department of Psychiatry and Dartmouth Institute, 1 Rope Ferry Road, Hanover NH, 03755, USA

[3]Medical Students, Geisel School of Medicine at Dartmouth College, 1 Rope Ferry Road, Hanover NH, 03755, USA

[4]Geisel School of Medicine at Dartmouth College, Department of Psychiatry, 1 Rope Ferry Road, Hanover NH, 03755, USA

[5]National Center for PTSD, 215 North Main Street, White River Junction, VT 05009, USA

[6]Veterans Rural Health Resource Center, White River Junction VA Medical Center, White River Junction, VT, 05009, USA

Abstract

Objective: There is limited knowledge about the ability of instruments to detect risk of suicide in a range of settings. Prior reviews have not considered whether the utility of instruments depends on prior probability of risk. We performed a systematic review to determine the diagnostic accuracy of instruments to detect risk of suicide in adults using likelihood ratio analysis. This method aids evaluation of prior probabilities of risk.

Data Sources: We searched Medline, Cochrane Database of Systematic Reviews, PsyINFO, EMBASE and Scopus from inception through January 19, 2021.

Study Selection: We included clinical trials, observational studies, and quasi-experimental studies assessing the diagnostic accuracy of instruments to detect risk of suicide in adults. There were no language restrictions.

Contact Information for Corresponding Authors: Natalie Riblet, MD MPH, Veterans Affairs Medical Center, Building 9, 215 North Main Street, White River Junction, VT, 05009. Natalie.Riblet@dartmouth.edu; Phone 1-802-295-9363; Fax: None.

Conflict of Interest Statement: The authors have no conflicts of interest to report.

Financial Disclosure: Natalie Riblet has no financial disclosures, Sarah Matsunaga has no financial disclosures, Angie Lee has no financial disclosures, Yinong Young-Xu has no financial disclosures, Brian Shiner has no financial disclosures, Paula Schnurr has no financial disclosures, Maxwell Levis has no financial disclosures, Bradley Watts has no financial disclosures.

Prior Presentations: None

Data Extraction: Three reviewers in duplicate assessed full texts to determine eligibility and extracted data from included studies. Positive (LR+) and negative likelihood ratio (LR–) and 95% Confidence Intervals were calculated for each instrument.

Results: Thirty studies met inclusion criteria. Most instruments showed minimal utility to detect or rule out risk of suicide with LR+ ≤ 2.0 and LR– ≥ 0.5. A few instruments had a high utility for improving risk detection in emergency room, inpatient mental health, and prison settings when patients score above the cutoff (LR+ > 10). For example, among patients discharged from an emergency room, the Columbia Suicide Severity Rating Scale – Clinical Practice Screener had a LR+ of 10.3 (95%CI: 6.3–16.8) at three-month follow-up. The clinical utility of the instruments depends on the pre-test probability of suicide in the setting. Because studies spanned over six decades, the findings are at risk for secular trends.

Discussion: We identified several instruments that may hold promise for detecting risk of suicide in emergency room, inpatient mental health, or prison settings. The utility of the instrument hinges, in part, on baseline suicide risk.

Registration: Prospero CRD42021285528

Keywords

Diagnostic accuracy; death by suicide; instruments; likelihood ratio analysis; systematic review; meta-analysis

Introduction

More than 20 self- or clinician-administered instruments have been developed to assist in detecting risk of suicide in adults.[1, 2] The Joint Commission has also set a standard that accredited organizations need to screen for suicidal ideation in patients evaluated (or treated) for behavioral health conditions using a validated instrument.[2] Available instruments generally have high face validity. A series of systematic reviews and meta-analysis, however, have concluded that there is a lack of robust evidence to support that these instruments can reliably detect risk of death by suicide. [3–7] In a meta-analysis of psychological scales, Runeson *et al.*[4] determined that no instrument met the authors' minimum criteria for diagnostic accuracy to detect risk of suicide (i.e. sensitivity >80% and specificity >50%).[4] Similarly, Carter *et al.* reported that the pooled positive predictive value (PPV) of psychological instruments to detect risk of suicide was only 5.5%.[7] In a 2013 systematic review completed for the U.S. Preventive Services Task Force, O'Connor concluded that there was only minimal evidence to support the routine practice of suicide screening in primary care settings.[8] Because instruments such as the Columbia Suicide Severity Rating Scale – Clinical Practice Screener (C-SSRS screener) continue to be routinely used in clinical practice,[2] it is necessary to determine which (if any) of these instruments can detect risk of suicide. Moreover, it is imperative that healthcare providers and policy makers are knowledgeable about how to use these instruments to inform suicide risk detection.

Historically, reviewers have evaluated the diagnostic accuracy of instruments to detect risk of suicide by analyzing instruments' predictive values as well as their sensitivity and specificity. According to Bayes' theorem, sensitivity and specificity can also be

simultaneously assessed using likelihood ratios and then combined with pretest probabilities to yield key insights about post-test probabilities.[9, 10] A likelihood ratio indicates how much more (or less) likely it is that a patient with suicide would have that test result as compared to a patient without suicide.[11] One can use the likelihood ratio for a given instrument to determine the applicability of the finding to their patient.[11] In other words, a healthcare provider or healthcare system can account for the prior probability of eventual suicide when applying an instrument to a patient, population or setting.[12] As such an instrument may perform better (or worse) in a particular setting based on prior knowledge of the characteristics of the population. This is a unique advantage of likelihood ratios. Positive and negative predictive values depend on the prevalence of risk in the sample.[11] While sensitivity and specificity assess an instrument's ability to predict the outcome, these values do not take into account the prior probability of risk.[11] Although meta-analyses based on likelihood ratio analysis have been successfully applied in other medical fields,[13, 14] this methodology has yet to be used to evaluate the diagnostic accuracy of instruments to detect risk of death by suicide.

The overall objective of this systematic review and meta-analysis is to fill this gap, evaluating the use of likelihood ratio analysis to evaluate the diagnostic accuracy of instruments to detect the risk of death by suicide in adults. To expand upon the current literature, we broadened our review to include the examination of instruments that were designed to assess risk of suicide regardless of underlying suicide risk or setting. The results of our review may uncover promising instruments to detect risk of suicide in various settings and motivate future research to design instruments with improved diagnostic accuracy. We chose death by suicide as the condition of interest because it is a societal goal to prevent death by suicide. Although intermediate outcomes such as suicidal ideation and non-fatal suicide attempts are more prevalent and therefore easier to measure in a study, these intermediary outcomes are far more susceptible to measurement bias.[15] These concerns about the measurement of intermediary suicide outcomes are very likely to be exacerbated when assessing the diagnostic accuracy of instruments.[16]

Methods

We conducted the review according to the PRISMA reporting guidelines for diagnostic test accuracy studies [17] and incorporated recommendations from the Cochrane Handbook for Diagnostic Test Accuracy Reviews.[18] The protocol is registered with Prospero (CRD42021285528).

Data Sources, Searches, Selection and Extraction

We searched Medline, Cochrane Database of Systematic Reviews (CDSR), PsyINFO, EMBASE and Scopus from inception through January 19, 2021. We used exploded MeSH terms and keywords to generate the following themes: psychological instruments, prediction, and suicidal behavior. We used the Boolean term "AND" to find the intersection between the three themes. We also reviewed the references of included studies.

We included randomized and non-randomized controlled trials as well as observational and quasi-experimental studies assessing the diagnostic accuracy of instruments to detect risk of

death by suicide in adult populations. We included studies that enrolled adult populations. If studies also enrolled non-adults, we used the following method to ensure only a limited number of subjects under age 18 were included in the sample. If age was reported as a continuous variable, we required that the mean (or median) age of the sample was 18 years or older. We reviewed the measures of variability to confirm that it was likely that patients younger than 18 years old accounted for a small proportion of the sample (i.e., <10%). If age was reported as a categorical variable (e.g., 15–19, 20–39), we reviewed the description of the sample (or reports of the underlying population) to confirm that it was likely that <10% of the sample was younger than 18 years. Because one study provided no patient-level characteristics, we contacted the hospital where the study was conducted to confirm that the psychiatric unit was an adult unit (i.e., ≥ 18 years) at the time of the study in 1970.[19] We also required that the instruments were clinician- or self-administered instruments that were designed with the primary intent to detect risk of suicide. We imposed no language restrictions.

We excluded studies that focused specifically on the diagnostic accuracy of instruments to detect risk of suicide in children and adolescents because this population is unique from adults. We also excluded any studies that reported insufficient data to evaluate the accuracy of the instrument.

Applying our *a priori* inclusion criteria, one reviewer (NR) screened the titles and abstracts of all potentially relevant studies, excluding those that were clearly ineligible. Three reviewers (NR, SM, YL) then independently and in duplicate assessed the full text of the remaining studies to determine eligibility. In the case of disagreement, a fourth reviewer (BW) independently evaluated these texts for inclusion. We used Rayyan software to facilitate the screening process.[20]

Three reviewers (NR, SM, YL) extracted data in duplicate from included studies. We extracted data related to demographics, methods, outcomes and risk of bias. We used the QUADAS 2.0 scale to evaluate risk of bias.[16] Discrepancies were resolved through consensus. This involved discussing the findings as a group and selecting the result that best described the data. Decisions about data selection were made irrespective of the seniority of the reviewer.

Data Synthesis

Because the instruments included in our review varied in their design and targeted divergent populations, we separately evaluated the diagnostic accuracy of each instrument. It was possible, however, that a study could contribute data to the analysis of more than one instrument.

For each instrument, we calculated the sensitivity, specificity, and corresponding positive (LR+) and negative likelihood ratio (LR−). We also calculated the 95% confidence intervals (CI).[21] In our study, a LR+ is a ratio of the chance of a positive response in the presence of suicide with the chances of a positive result in the absence of suicide. In this way, the LR+ tells you how much the probability of death by suicide increases based on the result. Conversely, the LR− is a ratio of the chance of a negative response in the absence of

suicide with the chances of a negative result in the presence of suicide. In this way, the LR- tells you how much the probability of death by suicide decreases based on the result. We conservatively applied a correction factor of 0.5 to cells in the case of zero values.[22, 23]

We observed that the Beck Suicide Intent Scale (SIS) and the Viennese Instrument for Suicidality in Correctional Institutions (VISCI) were each studied in four or more distinct populations. We applied bivariate mixed effects regression methods to calculate the summary estimates and summary receiver operator curves (SROC) for the SIS and the VISCI.[23, 24] We quantified heterogeneity that was due to threshold effects by examining the squared correlation coefficient which was calculated from the between-study covariance parameter.[23] In addition, we visually inspected the ROC plane.[23] Because studies of the VISCI included two distinct populations (pretrial; sentenced), we visually examined the data to assess for any trends suggestive of differences in outcomes based on the type of population. We sequentially removed studies to assess whether this resolved the observed variation and meaningfully changed our results.

We intended to apply a similar approach to the remaining instruments, but there were insufficient studies to permit bivariate analysis. We nonetheless felt that it was critical to present the individual results of each of these instruments as many of the instruments are used in clinical practice. Presenting the current evidence, as limited as it is, may help to inform future directions for research.

Data Analysis

We analyzed our results using a likelihood ratio analysis.[12, 22, 23] First, we created a graphical display of the LR+ and LR− of each instrument as well as the associated 95% CI. We defined clinical utility using the following approach: none (LR+ ≤ 2 or LR− ≥0.5), small (LR+ of >2 to ≤ 5 or LR− of 0.2 to < 0.5), moderate (LR+ of >5 to ≤ 10 or LR− of 0.1 to < 0.2), and high (LR+ of >10 or LR− 0.0 to <0.1).[25] A result was statistically meaningful if the CI stayed within clinical utility. We then generated a likelihood ratio scatterplot matrix.[22, 23] This matrix addresses concerns that separate pooling of LRs may overlook any correlation between the ratios.[18] Based on established, evidence-based criteria,[22, 23] results were assigned to one of four quadrants to further assess clinical utility: Right upper quadrant: (Exclusion & Confirmation, LR+ >10; LR− <0.1), Right lower quadrant: (Exclusion Only, LR+ ≤10, LR− <0.1), Left upper quadrant: (Confirmation Only, LR+ >10, LR− ≥0.1), and Left lower quadrant: (No Exclusion or Confirmation, LR+ ≤10, LR− ≥0.1).[22, 23] Because suicide is a low base-rate event, these cutoffs or greater are necessary in most populations to produce any clinically meaningful results.[12]

We used the MIDAS package in STATA 17 to perform the bivariate regression and generate the SROC.[23] All other quantitative analysis were performed using Microsoft Excel for Office 365.

If a study reported findings by subgroup (e.g., gender), we presented the results in this format because the authors frequently mentioned that they had observed differences in diagnostic accuracy based on these characteristics.

If the data for an instrument were collected but not reported in a format that would allow us to calculate sensitivity, specificity and corresponding likelihood ratios, we contacted the author for these data.[26–28] If we could not obtain the needed data from the authors,[26–28] we excluded the data-point on that particular instrument from our review. Specifically, we excluded results reported for the C-SSRS total[26] and certain subscales of the Columbia Classification System (C-CASA),[27] and the Manchester Self-Harm Rule (MSHR).[28]

Several studies reported estimates using different cutoff values.[26, 29–33] Here, we applied the following approach to select the data to present for our primary analysis. Based on prior guidance,[4] we gave first priority to results that yielded a sensitivity >80 % and a specificity > 50%.[4] We gave second priority to results that generated the highest sensitivity while maintaining a specificity around 50% or greater, and gave third priority to results that generated the highest sensitivity. To determine whether the choice of cutoff influenced our conclusions, we repeated the analysis using each alternative cutoff value. In addition, we observed that there was variability in follow-up time[26, 33–39] or choice of control.[40–42] We applied the same approach (as just described) to select the data to present for our primary analysis. We then repeated our analysis to see whether differences in these variables changed the results.

We used GRADEpro software[43] to evaluate the quality of the evidence for each instrument.[44] We rated quality as very low, low, moderate or high based on study design, risk of bias, indirectness, inconsistency, imprecision and publication bias.

Role of the Funding Source

NR has support from Department of Veterans Affairs Clinical Science Research & Development Career Development Award Program (MHBC-007-19F). ML is the recipient of a VA New England Early Career Development Award (VISN1 CDA-Levis). The supporters had no role in the analysis, interpretation, design, preparation, review or approval of this manuscript.

Results

We identified 11,547 potentially eligible records, of which 9,007 remained once we removed duplicates and ongoing studies (see Figure 1). After we applied our study inclusion and exclusion criteria to the potentially eligible records, we identified 41 reports (30 studies) that met study inclusion criteria.[19, 26–42, 45–67]

Table 1 summarizes the characteristics of one randomized trial, 22 cohort studies, and 7 case-control studies that met inclusion criteria. Studies represented a total of 31 instruments; some of which were modified versions of existing instruments. All studies were conducted in Europe or North America. The study years spanned from 1960 to 2018. Several studies specified age 18 years or older (or adult) as inclusion criteria or recruited subjects from an adult inpatient unit. There were only a few instruments that were tested in non-mental health populations or non-clinical settings. Several studies used a registry to identify suicide deaths. Eight studies included deaths that may have been misclassified as deaths due to undetermined causes, accidental poisoning or probable suicide.

Likelihood Ratio Analysis

As show in Figures 2 and 3, a small number of scales achieved a sensitivity > 80% and a specificity > 50%. Most instruments, however, had no utility for detecting risk of suicide when patients score about the cutoff or ruling out risk of suicide when patients score below the cutoff. There were just a few exceptions to these findings, listed as follows.

The C-SSRS screener, the Modified Screening for Suicide Risk of Prisoners (SSRP), and the Pallis 18-item + Beck Suicide Intent Scale (SIS) 7-item had a high utility (LR+ 10+) for detecting risk of suicide when patients score above the cutoff (Figure 2). While the Suicide Potential Scale (SPS) had a high utility, the CI included no utility (LR+ 12, 95%CI: 1.8–81.7). The VISCI was the only instrument that had a small utility for ruling out risk of suicide when patients score below the cutoff and results did not cross no utility (LR− 0.2, 95%CI: 0.17–0.3).

Related to these findings, the likelihood ratio matrix found that most instruments were not useful for detecting or ruling out risk of suicide (see Figure 4). There were, however, a few exceptions. The C-SSRS screener had high utility in an emergency room population for detecting risk of suicide at three-month follow-up when patients had a positive screen (i.e., 'yes' to any of three questions related to intensity of suicidal thoughts and history of self-harm) (LR+ 10.3, LR− 0.7). In a case-control study, the Modified SSRP also had high utility for detecting risk of suicide among pre-trial inmates when patients scored 3+ (LR+ 10.5, LR− 0.3). In addition, at 12-month follow-up, the Pallis 18-item + Beck SIS 7-item had high utility in an inpatient or emergency room sample for detecting risk of suicide when patients scored g+ (LR+ 10.1) and bordered on high utility to rule out risk of suicide when patients scored below the cutoff (LR− 0.1). Finally, the Beck SIS 4-item bordered on high utility for ruling out risk of suicide when patients scored below 6 (LR+ 2.3, LR− 0.1).

Sensitivity Analysis

The summary estimates for the five studies of the Beck SIS Total had threshold effects upwards of 100%. There was notable variation in populations, selected cutoff values and follow-up time across studies. Although we were unable to resolve these threshold effects, the results remained unchanged regardless of the combination of studies. The summary estimates for VISCI had threshold effects upwards of 100% and had no utility for detecting risk of suicide. Because this estimate included two distinct populations (sentenced; pre-trial), we analyzed the results that the authors reported for each population. We found that the VISCI had utility for detecting risk of suicide in sentenced populations based on high positive LRs in the validation (e.g., LR+ 13) and index sample (e.g., LR+ 38). The VISCI, however, had only small to moderate utility for detecting risk of suicide in pretrial populations in the validation sample.

Several studies reported multiple cutoff points and/or follow-up periods. In general, we found that the choice of cutoff or follow-up did not change our conclusions. However, in the case of the Pallis 18-item + Beck SIS 7-item, the high utility fell to moderate at 6- and 24 – month follow-up (LR+ 9.0; LR− 0.1–0.2). The C-SSRS screener had moderate utility for detecting risk of suicide among those scoring above the cutoff at one- (LR+ 5.0, LR− 0.8),

six- (LR+ 7.3, LR− 0.8), and 12- month follow-up (LR+ 7.5, LR− 0.8). Finally, using an index sample, Motto *et al.* (1976) reported an extremely high LR+ >50 for the Motto Risk Estimate for detecting risk of suicide at 12-month follow-up among hospitalized patients.[37] The authors were unable to replicate this result in a validation sample at 12-month follow-up (LR+ 2.3) or in a larger sample at 24-months (LR+2.5).[36–38]

Quality Assessment

As shown in Supplementary Table 1, all studies were judged to be at low risk when considering applicability of patient selection, index test and reference standard. With regards to risk of bias, most studies were also at low risk of bias for patient selection, reference standard and flow and timing. There were some concerns, however, that there was insufficient information provided in the majority of studies to judge the administration of the index test. The degree to which bias may have influenced the findings of case-control studies was also, generally, less clear.

According to the GRADE analysis, the quality of evidence for most instruments was high (i.e., Beck SIS 4-item, Beck SSI current and worst, C-CASA, MSHR, ReACT, and Pallis 6 plus Beck SIS 7-item) or moderate (i.e., Beck SIS 7-item, Buglass and Horton 3- and 6-item, C-SSRS screener, Neuropsychiatric Hospital Suicide Prediction Schedule, Pallis 18 plus Beck SIS 7-item, Pierce, The suicidal risk assessment scale of Ducher (RSD), Modified SAD PERSONS Scale, South London and Maudsley NHS Foundation Trust (SLaM), SPS, Revised SPS, SSRP, Modified SSRP, Suicide Probability Scale, and VISCI). The results were downgraded in some cases due to wide confidence intervals or inconsistency. There was no evidence of publication bias.

Discussion

Over the past 60 years, studies have reported on a number of instruments that are designed to detect risk of suicide in diverse populations, ranging from those with a known history of suicidal behavior to those with any acute health symptoms. The instruments have been applied in divergent settings. Aligned with the literature, [3–7] our review has concluded that most instruments show minimal utility to detect or rule out risk of suicide. A few instruments that may hold promise in improving the ability to detect risk of suicide include the C-SSRS screener in emergency room settings, the Modified SSRP for pre-trial inmates, the VISCI for pre-trial and sentenced inmates, and the Pallis 18-item + Beck SIS 7-item and the Beck SIS 4-item in acute psychiatric settings.

Simpson *et al.* examined the C-SSRS screener in 92,643 patients who presented to an emergency room with any acute health concern,[39] While the authors identified that a positive screen had a high utility for detecting risk of suicide at 90-day follow-up (i.e., LR+ 10.3), the instrument performed poorly at ruling out cases (i.e., LR− 0.7). This is not surprising given the instrument's poor sensitivity of 37. Practically speaking, the clinical relevance of their findings also remains unclear. The incidence of suicide in their population at 90-days was 0.03%, meaning that a robust LR+ of 10.3 had virtually no effect on modifying a patient's post-test probability of suicide (i.e., 0.3%). It is conceivable that the screen may be useful when applied to population with a higher prevalence of suicide. For example,

Geulayov *et al.* (2019) reported that the probability of suicide within the first three months of emergency room discharge among patients who presented with suicidal behavior was greater than 0.65%.[68] This means that a LR+ of 10.3 would increase the post-test probability of suicide to 6.3%, a potentially meaningful finding. Yet, because the C-SSRS screener asks about suicidal ideation and self-harm, most (if not all) of these patients would screen positive on the instrument, nullifying any possible benefit. This suggests that the C-SSRS screener may only be useful in a population of patients whose baseline suicide risk is high (e.g., alcohol use disorder),[69] but the chief complaint is not suicidal behavior. Of course, future research would need to confirm this.

Among 30 pre-trial detainees who died by suicide, Dahle *et al.* (2005) determined that the modified SSRP had high utility for detecting risk of suicide (LR+ 10.5) and small utility for ruling out risk of suicide (LR− 0.3).[66] It is important to highlight that the likelihood ratio analysis uncovered a potential application for the modified SSRP to improve the ability to detect risk of suicide, even though the sensitivity of the instrument was relatively low (i.e., 70). Frottier *et al.* (2008; 2009) also concluded that among 228 inmates who died by suicide, the VISCI had a high utility for detecting risk of suicide among sentenced inmates with LR+s ranging from 13 to 38.[30, 31] Notably, the VISCI was the only instrument that had a small utility for ruling out risk of suicide (LR− 0.2) and results did not cross no utility. Given the high rates of suicide in inmates,[70–72] both instruments may have utility in improving the detection of risk of suicide in real-world practice. It remains unclear whether these instruments could produce similar results in other high-risk populations. Inmates have unique risk factors for suicide [72] and the instruments include several items about legal concerns. The studies also used a case-control design.

We determined that a combined instrument (Pallis 18-item + Beck SIS 7-item) had high utility in an inpatient or emergency room sample for improving the detection of risk of suicide (LR+ 10.1) and bordered on high utility to rule out risk of suicide (LR− 0.1) when patients score above the cutoff.[33] Considering that the rates of suicide after psychiatric hospitalization are much higher than the general population,[73] this may be a clinically meaningful finding. For example, if the pre-test probability of suicide in the first three months after psychiatric discharge is 1.1%,[73] then the instrument would increase the post-test probability of suicide to approximately 10% among those scoring above the cutoff. It also means that if a patient scores below the cutoff, then the probability of suicide is exceedingly small (i.e., 0.1%.). While this result is quite promising, several factors must be considered. By design, the Beck SIS can only be administered to a patient with a current (or prior) history of suicide attempt and therefore, has limited application. The instruments were studied in patients who presented to a hospital or emergency room setting. The findings were also most evident at 12-month follow-up and became less pronounced at 6- and 24-month follow-up. Finally, Pallis *et al.* (1984)[33] conducted their study more than 35 years ago and the rates of suicide in the population have shifted over this timeframe.[74, 75, 76] Interventions to manage suicide risk in high-risk populations have also evolved.[77, 78] These factors could influence suicide risk post-discharge.

Aligned with prior reviews, we noted that most instruments had negligible utility to detect or rule out risk of death by suicide. For example, we found no evidence to support that

the SAD PERSONS scale or modified SAD PERSONS scale improved the ability to detect risk of suicide. In their review, Runeson *et al.* (2017) also concluded that these scales had low diagnostic accuracy to detect risk of suicide.[4] Notably, Runeson *et al.* recommended that the SAD PERSONS scale and its modified version should not be used in their current format.[4] We also determined that decision rules had minimal utility to detect or rule out risk of suicide. Initially, this result may seem surprising because decision rules have been found to have high sensitivity. It is important to point out, however, that decision rules perform poorly at detecting patients who are at low risk of suicide. As an example, in a study of the MSHR in patients who presented with self-harm to an emergency room, Steeg *et al.* (2018) determined that the sensitivity of the MHSR was 89, while the specificity was only 11.[62]

Overall, we noted that there was a fair amount of overlap in the types of items that were described in the instruments. It was common for scales to include items that assessed for suicidal behavior, mental health symptoms or clinical-demographic information. In addition, while there is emerging evidence to suggest that combining suicide scales with machine learning may improve the detection of risk of suicidal behavior,[79] none of the included studies employed these methods.

Strengths and Limitations of the Review

Our review has several strengths. We applied a systematic approach to identify studies and applied no language restrictions. We covered a broad range of instruments including several not previously discussed. Our decision to focus on death by suicide mitigated concerns for measurement bias and reassuringly, several studies included deaths due to undetermined cause or accidental poisoning.[80] Our use of likelihood ratio analysis may assist providers and researchers in clarifying the applicability of an instrument in a given context based on population or setting.

There are limitations to our review. First, we are unable to comment on the role of instruments in detecting suicide risk in regions outside of North America and Europe. Because included instruments were studied over a span of nearly six decades in several countries. For example, discharge care for patients admitted to a mental health unit in the 1960s may look very different from those of the 2000s. Second, while it was useful that studies tended to follow patients for a long period of time, this poses the risk that other variables could better explain our findings. Third, few studies evaluated short or near-term risk of death by suicide. This is a critical gap in knowledge as certain healthcare professionals such as an emergency room personnel may have little or no contact with a patient during any given year. Therefore, an emergency room clinician would benefit more knowing about the patient's near-term risk in the next 30 days or 60 days versus the next two or three years. Fourth, we did not identify any instrument that has examined suicide risk detection in primary care settings. Yet, instruments such as the C-SSRS are frequently implemented in these settings.[2] While two reviewers assessed the full text of studies to determine eligibility, only one reviewer performed the title and abstract step. Therefore, it is possible that we may have missed additional studies that met our inclusion criteria. Fifth, while our review identified many instruments to detect risk of suicide, the instruments were typically examined in only one or two studies. In the two cases where we

were able to perform bivariate analysis, we noted a large amount of heterogeneity due to threshold effects. We were unable to resolve the heterogeneity. A few studies also included younger patients (usually ≥ 15 years). It is unlikely, however, that the inclusion of these patients biased our results because these patients accounted for a small proportion of the study samples. Finally, a handful of instruments (e.g. RSD[48]) met the minimum criteria for diagnostic accuracy to detect risk of suicide based on sensitivity and specificity[4] but did not have high utility (LR+ >10) for detecting risk of suicide. In these cases, studies usually had small samples and zero cells, limiting the interpretation of the results. Nonetheless, it may be useful to study these instruments more rigorously.

In summary, the evidence in support of the use of any instrument to detect risk of suicide is limited. While we did not identify any instruments that are useful for detecting risk of suicide in primary care or specialty medical settings, we located several scales that may hold promise in other settings. Specifically, the C-SSRS screener may be useful in emergency room settings to screen patients who are at high risk of suicide but whose presenting symptom is not suicidal behavior. Conversely, the Modified SSRP or VISCI may be beneficial in incarcerated populations, and the Pallis 18-item + Beck Suicide Intent Scale (SIS) 7-item may be helpful in psychiatrically hospitalized patients. Because these suggestions are based on limited evidence, it is important that future research further study these promising instruments. Our work also highlights the importance of selecting the correct instrument for a specific situation as the setting, population and follow-up time are important considerations. Ultimately, there is a need for researchers not only to study instruments to detect risk of suicide in other settings (e.g., primary care), but also to develop new and better ways for providers to detect risk of suicide in patients in real-time.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

No additional individuals contributed to this work. The views expressed in this article do not necessarily represent the views of the Department of Veterans Affairs or of the United States government. No financial disclosures were reported by the authors of this paper.

Grant Funding:

Dr. Riblet has support from the Department of Veterans Affairs Clinical Science Research & Development Career Development Award Program (MHBC-007-19F). Dr. Levis is the recipient of a VA New England Early Career Development Award (VISN1 CDA-Levis). The supporters had no role in the design, analysis, interpretation, or publication of this study.

References

1. Brown G A review of suicide assessment measures for intervention research with adults and older adults. https://www.sprc.org/resources-programs/review-suicide-assessment-measures-intervention-research-adults-older-adults Accessed December 9, 2021.
2. The Joint Commission. 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 December 9, 2021

3. Large M, Kaneson M, Myles N, et al. Meta-Analysis of longitudinal cohort studies of suicide risk assessment among psychiatric patients: Heterogeneity in results and lack of improvement over time. PLoS One. 2016;11(6):e0156322. [PubMed: 27285387]
4. Runeson B, Odeberg J, Pettersson A, et al. Instruments for the assessment of suicide risk: A systematic review evaluating the certainty of the evidence. PLoS One. 2017;12(7):e0180292. [PubMed: 28723978]
5. Saab MM, Murphy M, Meehan E, et al. Suicide and self-harm risk assessment: A systematic review of prospective research. Arch Suicide Res. 2021:1–21.
6. Chan MK, Bhatti H, Meader N, et al. Predicting suicide following self-harm: systematic review of risk factors and risk scales. Br J Psychiatry. 2016;209(4):277–83. [PubMed: 27340111]
7. Carter G, Milner A, McGill K, et al. Predicting suicidal behaviours using clinical instruments: systematic review and meta-analysis of positive predictive values for risk scales. Br J Psychiatry. 2017;210(6):387–95. [PubMed: 28302700]
8. O'Connor E, Gaynes BN, Burda BU, Soh C, Whitlock EP. Screening for and treatment of suicide risk relevant to primary care: a systematic review for the U.S. Preventive Services Task Force. Ann Intern Med. 2013;158(10):741–54. [PubMed: 23609101]
9. Johnson KM. Using Bayes' rule in diagnostic testing: a graphical explanation. Diagnosis. 2017;4(3):159–67. [PubMed: 29536931]
10. Johnson KM. Erratum to: Using Bayes' rule in diagnostic testing: a graphical explanation. Diagnosis. 2018;5(2):89-. [PubMed: 29813027]
11. Deeks JJ, Altman DG. Diagnostic test 4: Likelihood ratios. BMJ. 2004; 329: 168–169. [PubMed: 15258077]
12. Aitken C, Mavridis D. Reasoning under uncertainty. Evidence Based Mental Health. 2019;22:44–8. [PubMed: 30679196]
13. Cai W, Zhu D, Byanju S, et al. Magnetic resonance spectroscopy imaging in diagnosis of suspicious prostate cancer: A meta-analysis. Medicine. 2019;98(14):e14891. [PubMed: 30946315]
14. Downie A, Williams CM, Henschke N, et al. Red flags to screen for malignancy and fracture in patients with low back pain: Systematic review. BMJ. 2013;347:f7095. [PubMed: 24335669]
15. Glenn C, Nock M. Improving the short-term prediction of suicidal behavior. Am J Prev Med. 2014;47(3S2):S176–S80. [PubMed: 25145736]
16. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36. [PubMed: 22007046]
17. McInnes MDF, Moher D, Thombs BD et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018;319(4):388–396 [PubMed: 29362800]
18. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. https://methods.cochrane.org/sdt/handbook-dta-reviews Accessed December 9, 2021.
19. van de Loo KJ, Diekstra RW. The construction of a questionnaire for the prediction of subsequent suicidal attempts: a preliminary study. Nederlands tijdschrift voor de psychologie en haar grensgebieden. 1970;25:95–100. [PubMed: 5417304]
20. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan - a web and mobile app for systematic reviews. Systematic Reviews. 2016; 5:210, DOI: 10.1186/s13643-016-0384-4. 2016. [PubMed: 27919275]
21. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J Clin Epidemiol. 1991;44(8):763–70. [PubMed: 1941027]
22. Stengel D, Bauwens K, Sehouli J, Ekkernkamp A, Porzsolt F. A likelihood ratio approach to meta-analysis of diagnostic studies. J Med Screen. 2003;10(1):47–51. [PubMed: 12790315]
23. Dwamena B, Sylvester R, Carlos RC. MIDAS: meta-analysis of diagnostic accuracy studies. http://fmwww.bc.edu/repec/bocode/m/midas.pdf Accessed October 21, 2021.

24. Reitsma JB, Glass ASG, Rutjes AWS, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58(10):982–90. [PubMed: 16168343]
25. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article abuot diagnostic test. B. What are the results and will they help me in caring for my patients? JAMA. 1994;271(9):703–7. [PubMed: 8309035]
26. Lindh ÅU, Dahlin M, Beckman K, et al. A comparison of suicide risk scales in predicting repeat suicide attempt and suicide: A clinical cohort study. J Clin Psychiat. 2019;80(6).
27. Randall JR, Sareen J, Chateau D, Bolton JM. Predicting future suicide: Clinician opinion versus a standardized assessment tool. Suicide Life Threat Behav. 2019;49(4):941–51. [PubMed: 29920749]
28. Steeg S, Kapur N, Webb R, et al. The development of a population-level clinical screening tool for self-harm repetition and suicide: the ReACT Self-Harm Rule. Psychol Med. 2012;42(11):2383–94. [PubMed: 22394511]
29. Buglass D, Horton J. A scale for predicting subsequent suicidal behaviour. Br J Psychiatry. 1974;124:573–8. [PubMed: 4850288]
30. Frottier P, Koenig F, Matschnig T, Seyringer M-E, Frühwald S. [VISCI- Viennese Instrument for Suicidality in Correctional Institutions]. Psychiat Prax. 2008;35:e1–e8.
31. Frottier P, Koenig F, Seyringer M, Matschnig T. The distillation of "VISCI": Towards a better identification of suicidal inmates. Suicide Life Threat Behav. 2009;39(4):376–85. [PubMed: 19792979]
32. Kuerz A, Moeller HJ, Torhorst A, Lauter H. Validation of six risk scales for suicide attempters. Current Issues of Suicidology. 1988:174–7.
33. Pallis DJ, Gibbons JS, Pierce DW. Estimating suicide risk among attempted suicides: II Efficiency of predictive scales after the attempt. Br J Psychiatry. 1984;144:139–48. [PubMed: 6704599]
34. Buglass D, McCulloch JW. Further suicidal behaviour: The development and validation of predictive scales. Br J Psychiatry. 1970;116:483–91. [PubMed: 5460558]
35. Katz C, Randall JR, Sareen J, et al. Predicting suicide with the SAD PERSONS scale. Depress Anxiety. 2017;34(9):809–16.36. [PubMed: 28471534]
36. Motto JA. Preliminary field-testing of a risk estimator for suicide. Suicide Life Threat Behav. 1985;15(3):139–50. [PubMed: 4089933]
37. Motto JA, Heilbron DC. Development and validation of scales for estimation of suicidal risk. Suicidology. 1976:166–99.
38. Motto JA, Heilbron DC, Juster RP. Development of a clinical instrument to estimate suicide risk. Am J Psychiat. 1985;142(6):680–6. [PubMed: 4003585]
39. Simpson S, Goans C, Loh R, et al. Suicidal ideation is insensitive to suicide risk after ED discharge: performance characteristics of the Columbia-Suicide Severity Rating Scale Screener. Acad Emerg Med. 2021; 28(6): 621–629. [PubMed: 33346922]
40. Braucht GN, Wilson LT. Predictive utility of the revised suicide potential scale. J Consult Clin Psych. 1970;35(3):426.
41. Dean RA, Miskimins WM, De Cook R, Wilson LT, Maley RF. Prediction of suicide in a psychiatric hospital. J Clin Psychol. 1967:296–301. [PubMed: 6082115]
42. Wang AG, Nielsen B, Bille-Brahe U, Hansen W, Kolmos L. Attempted suicide in Denmark. III. Assessment of repeated suicidal behaviour. Acta Psychiatr Scand. 1985;72(4):389–94. [PubMed: 4072741]
43. GRADEpro GDT: GRADEpro Guideline Development Tool [Software]. McMaster University and Evidence Prime, 2021. gradepro.org. Accessed December 14, 2021;
44. Schünemann H, Brożek J, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. The GRADE Working Group; 2013. guidelinedevelopment.org/handbook. Accessed December 14, 2021.
45. Beck AT, Brown GK, Steer RA, Dahlsgaard KK, Grisham JR. Suicide ideation at its worst point: A predictor of eventual suicide in psychiatric outpatients. Suicide and Life-Threat Behav. 1999;29(1):1–9.

46. Brown GK, Beck AT, Steer RA, et al. Risk factors for suicide in psychiatric outpatients: A 20-year prospective study. J Consult Clin Psych. 2000;68(3):371–7.
47. Joiner TE Jr., Steer RA, Brown G, et al. Worst-point suicidal plans: A dimension of suicidality predictive of past suicide attempts and eventual death by suicide. Behav Res Ther. 2003;41(12):1469–80. [PubMed: 14583414]
48. Ducher JL, Terra JL. [Is the suicidal risk assessment scale RSD of predictive value?]. LEncphale. 2006;32(5Pt1):738–45.
49. Clark DC, Young MA, Scheftner WA, Fawcett J, Fogg LA field test of Motto's Risk Estimator for Suicide. Am J Psychiat. 1987;144(7):923–6. [PubMed: 3605405]
50. Cooper J, Kapur N, Dunning J, et al. A clinical tool for assessing risk after self-harm. Ann Emerg Med. 2006;48(4):459–66. [PubMed: 16997684]
51. Harriss L, Hawton K. Suicidal intent in deliberate self-harm and the risk of suicide: The predictive power of the Suicide Intent Scale. J Affect Disorders. 2005;86(2–3):225–33. [PubMed: 15935242]
52. Harriss L, Hawton K, Zahl D. Value of measuring suicidal intent in the assessment of people attending hospital following self-poisoning or self-injury. Br J Psychiatry. 2005;186(1):60–6. [PubMed: 15630125]
53. Lindh AU, Waern M, Beckman K, et al. Short term risk of non-fatal and fatal suicidal behaviours: The predictive validity of the Columbia-Suicide Severity Rating Scale in a Swedish adult psychiatric population with a recent episode of self-harm. BMC Psychiatry. 2018;18(1):319. [PubMed: 30285661]
54. Lindh ÅU, Beckman K, Carlborg A, et al. Predicting suicide: A comparison between clinical suicide risk assessment and the Suicide Intent Scale. J Affect Disorders. 2020;263:445–9. [PubMed: 31969276]
55. Lopez-Morinigo JD, Fernandes AC, Shetty H, et al. Can risk assessment predict suicide in secondary mental healthcare? Findings from the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register. Soc Psychiatry Psychiatr Epidemiol. 2018;53(11):1161–71. [PubMed: 29860569]
56. Naud H, Daigle MS. Predictive validity of the suicide probability scale in a male inmate population. J Psychopathol Behav Assess. 2010;32:333–42.
57. Niméus A, Alsén M, Träskman-Bendz L. The Suicide Assessment Scale: An instrument assessing suicide risk of suicide attempters. European Psychiatry. 2000;15(7):416–23. [PubMed: 11112934]
58. Skogman K, Alsén M, Ojehagen A. Sex differences in risk factors for suicide after attempted suicide-a follow-up study of 1052 suicide attempters. Soc Psychiatry Psychiatr Epidemiol. 2004;39(2):113–20. [PubMed: 15052392]
59. Niméus A, Alsén M, Träskman-Bendz L. High suicidal intent scores indicate future suicide. Archives of Suicide Research. 2002;6(3):211–9.
60. Probert-Lindström S, Berge J, Westrin Å, Ojehagen A, Pavulans KS. Long-term risk factors for suicide in suicide attempters examined at a medical emergency in patient unit: results from a 32-year follow-up study. BMJ Open. 2020;10(10):e038794.
61. Samuelsson M, Jokinen J, Nordstrom AL, Nordstrom P. CSF 5-H1AA, suicide intent and hopelessness in the prediction of early suicide in male high-risk suicide attempters. Acta Psychiatr Scand. 2005;113:44–7.
62. Steeg S, Quinlivan L, Nowl R, et al. Accuracy of risk scales for predicting repeat self-harm and suicide: A multicentre, population-level cohort study using routine clinical data. BMC Psychiatry. 2018;18(1): 113. [PubMed: 29699523]
63. Stefansson J, Nordström P, Jokinen J. Suicide Intent Scale in the prediction of suicide. J Affect Disorders. 2012;136(1–2):161–71.
64. Stefansson J, Nordstrom P, Jokinen J. Suicide intent scale in the prediction of suicide. Eur Psychiat. 2011;26. P03–479.
65. Stefansson J, Nordström P, Runeson B, Asberg M, Jokinen J. Combining the Suicide Intent Scale and the Karolinska Interpersonal Violence Scale in suicide risk assessments. BMC Psychiatry. 2015;15: 226. [PubMed: 26400812]

66. Dahle K-P, Lohner JC, Konrad N. Suicide prevention in penal institutions: validation and optimization of a screening tool for early identification of high-risk inmates in pretrial detention. Int J Forensic Ment. 2005;4(1):53–62.

67. Farberow NL, MacKinnon D. A suicide prediction schedule for neuropsychiatric hospital patients. J Nerv Ment Dis. 1974;158(6):408–19. [PubMed: 4841286]

68. Geulayov G, Casey D, Bale L, et al. Suicide following presentation to hospital for non-fatal self-harm in the Multicentre Study of Self-harm: a long-term follow-up study. Lancet Psychiat. 2019;6(12):1021–30.

69. Brådvik L Suicide Risk and Mental Disorders. Int J Env Res Pub He. 2018;15(9):2028.

70. Kajeepeta S, Mauro PM, Keyes KM, et al. Association between county jail incarceration and cause-specific county mortality in the USA, 1987–2017: a retrospective, longitudinal study. Lancet Public Health. 2021;6(4):e240–e8. [PubMed: 33636104]

71. Bukten A, Stavseth MR. Suicide in prison and after release: a 17-year national cohort study. Eur J Epidemiol. 2021;36(10):1075–83. [PubMed: 34427828]

72. Zhong S, Senior M, Yu R, et al. Risk factors for suicide in prisons: a systematic review and meta-analysis. Lancet Public Health. 2021;6(3):e164–e74. [PubMed: 33577780]

73. Chung DT, Ryan CJ, Hadzi-Pavlovic D, et al. Suicide rates after discharge from psychiatric facilities: A systematic review and meta-analysis. JAMA Psychiatry. 2017;74(7):694–702. [PubMed: 28564699]

74. Hedegaard H, Curtin SC, Warner M. Increase in suicide mortality in the United States, 1999–2018. NCHS Data Brief. 2020(362):1–8.

75. Global Naghavi M., regional, and national burden of suicide mortality 1990 to 2016: systematic analysis for the Global Burden of Disease Study 2016. BMJ. 2019;364:l94. [PubMed: 31339847]

76. McKeown RE, Cuffe SP, Schulz RM. US suicide rates by age group, 1970–2002: an examination of recent trends. Am J Public Health. 2006;96(10):1744–51. [PubMed: 17008567]

77. Mann JJ, Michel CA, Auerbach RP. Improving suicide prevention through evidence-based strategies: A Systematic review. Am J Psychiatry. 2021;178(7):611–24. [PubMed: 33596680]

78. Riblet NBV, Shiner B, Young-Xu Y, Watts BV. Strategies to prevent death by suicide: meta-analysis of randomised controlled trials. Br J Psychiatry. 2017;210(6):396–402. [PubMed: 28428338]

79. Haroz EE, Kitchen C, Nestadt PS, et al. Comparing the predictive value of screening to the use of electronic health record data for detecting future suicidal thoughts and behavior in an urban pediatric emergency department: A preliminary analysis. Suicide Life Threat Behav. Dec 2021;51(6):1189–1202. [PubMed: 34515351]

80. Snowdon J, Choi NG. Undercounting of suicides: Where suicide data lie hidden. Glob Public Health. 2020;15(12):1894–901. [PubMed: 32744898]

Clinical Points

- The utility of scales to detect risk of suicide may relate to prior probability of risk but reviews have not studied this.
- There is limited evidence to demonstrate that most scales can detect suicide, and promising scales require further study.
- In some settings (e.g., psychiatric hospital), there may be scales that can improve risk detection based on higher, post-test probability.

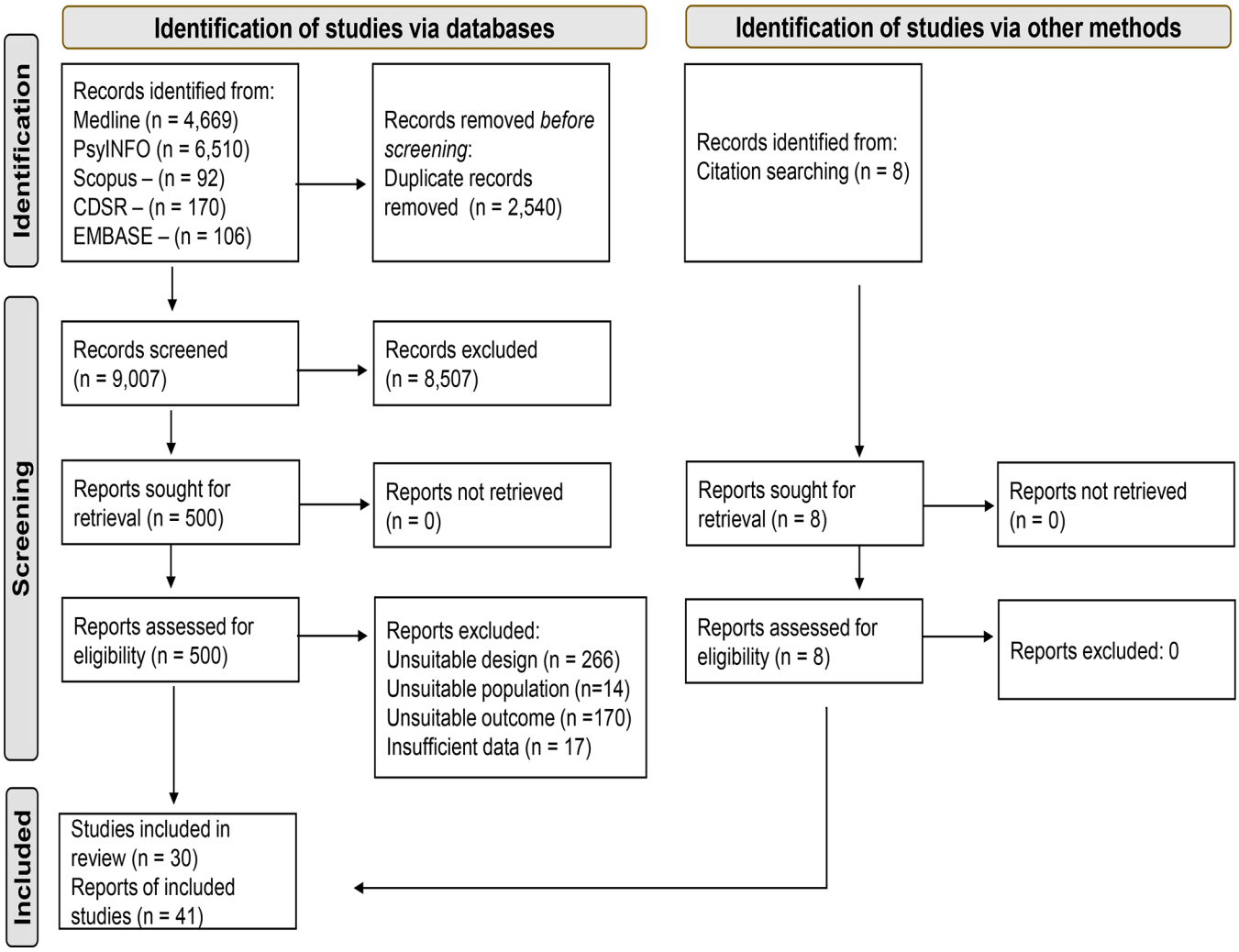

Figure 1:
Prisma Flow Diagram

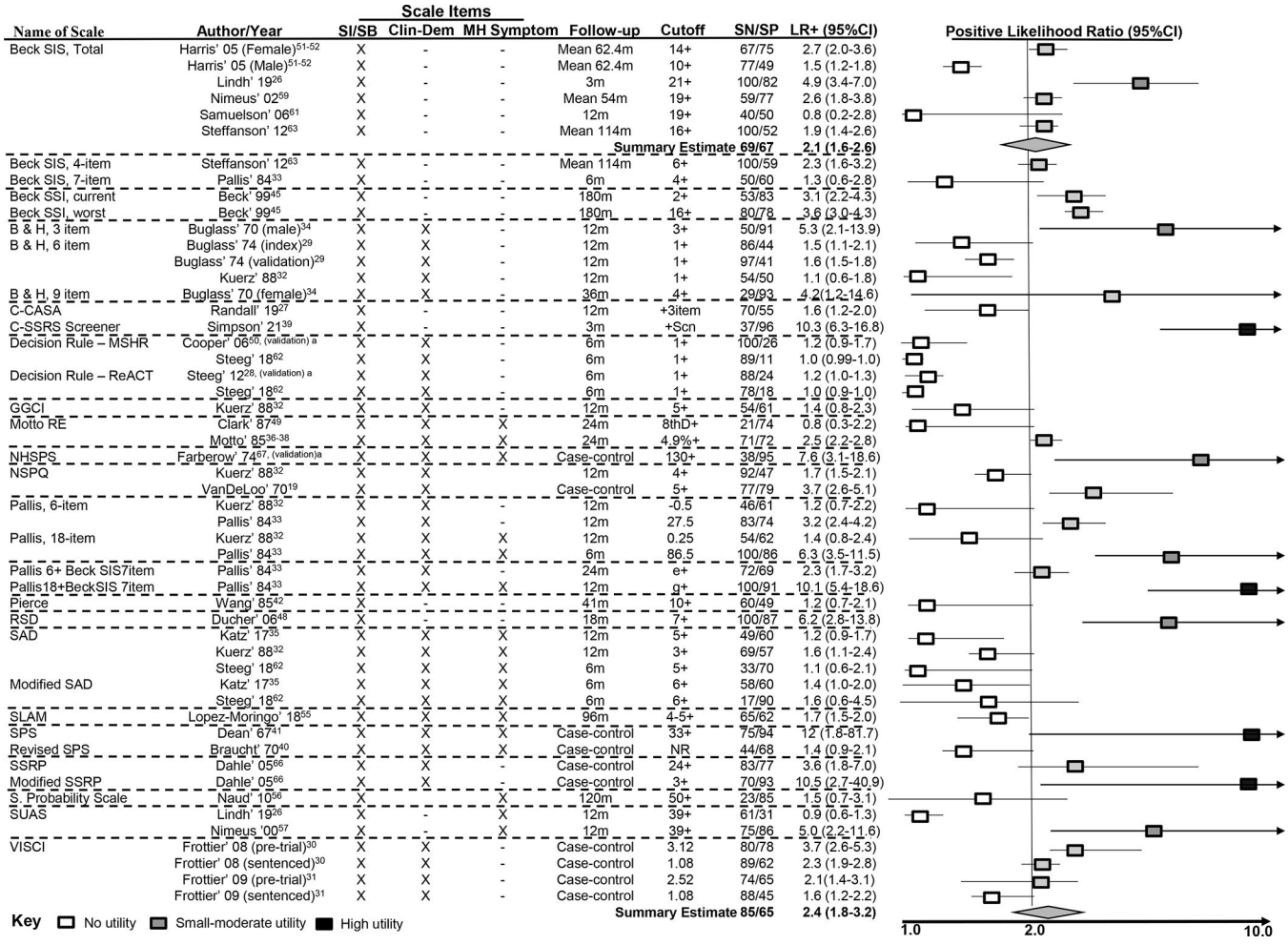

Figure 2:

Ability of instruments to detect risk of suicide when patients score at or above the cutoff. CI = Confidence Interval; Clin-Dem = Clinical and/or demographic characteristics; m = months; MH = Mental health; SI/SB = Suicidal ideation and/or suicidal behavior; SN = Sensitivity; SP = Specificity; LR = Likelihood Ratio;

[a] The studies found that LR+ were similar across derivation and validation sets.

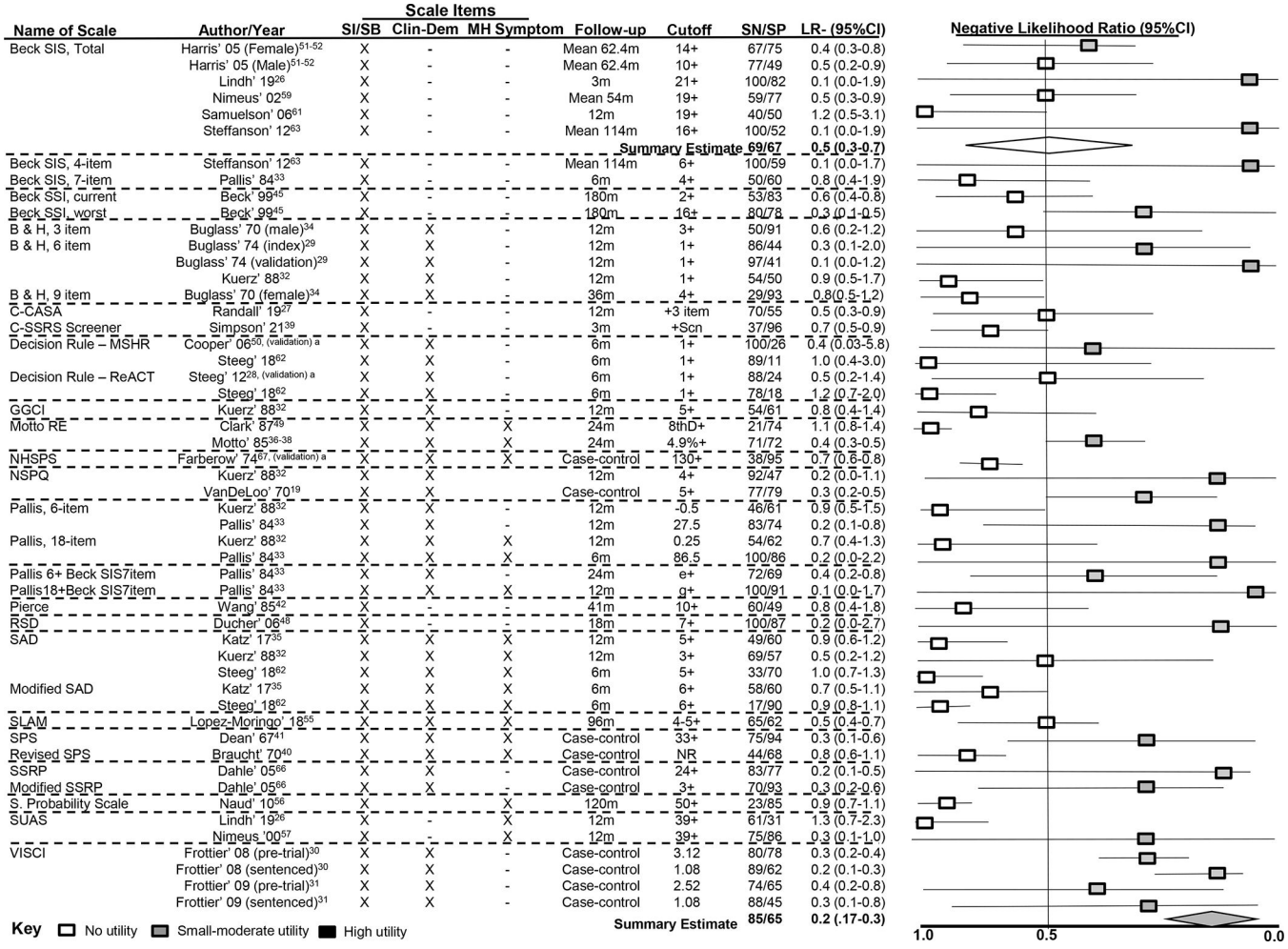

Figure 3:

Ability of instruments to rule out risk of suicide when patients score below the cutoff. CI = Confidence Interval; Clin-Dem = Clinical and/or demographic characteristics; m = months; MH = Mental health; SI/SB = Suicidal ideation and/or suicidal behavior; SN = Sensitivity; SP = Specificity; LR = Likelihood Ratio;

[a] The studies found that LR− were similar across derivation and validation sets.

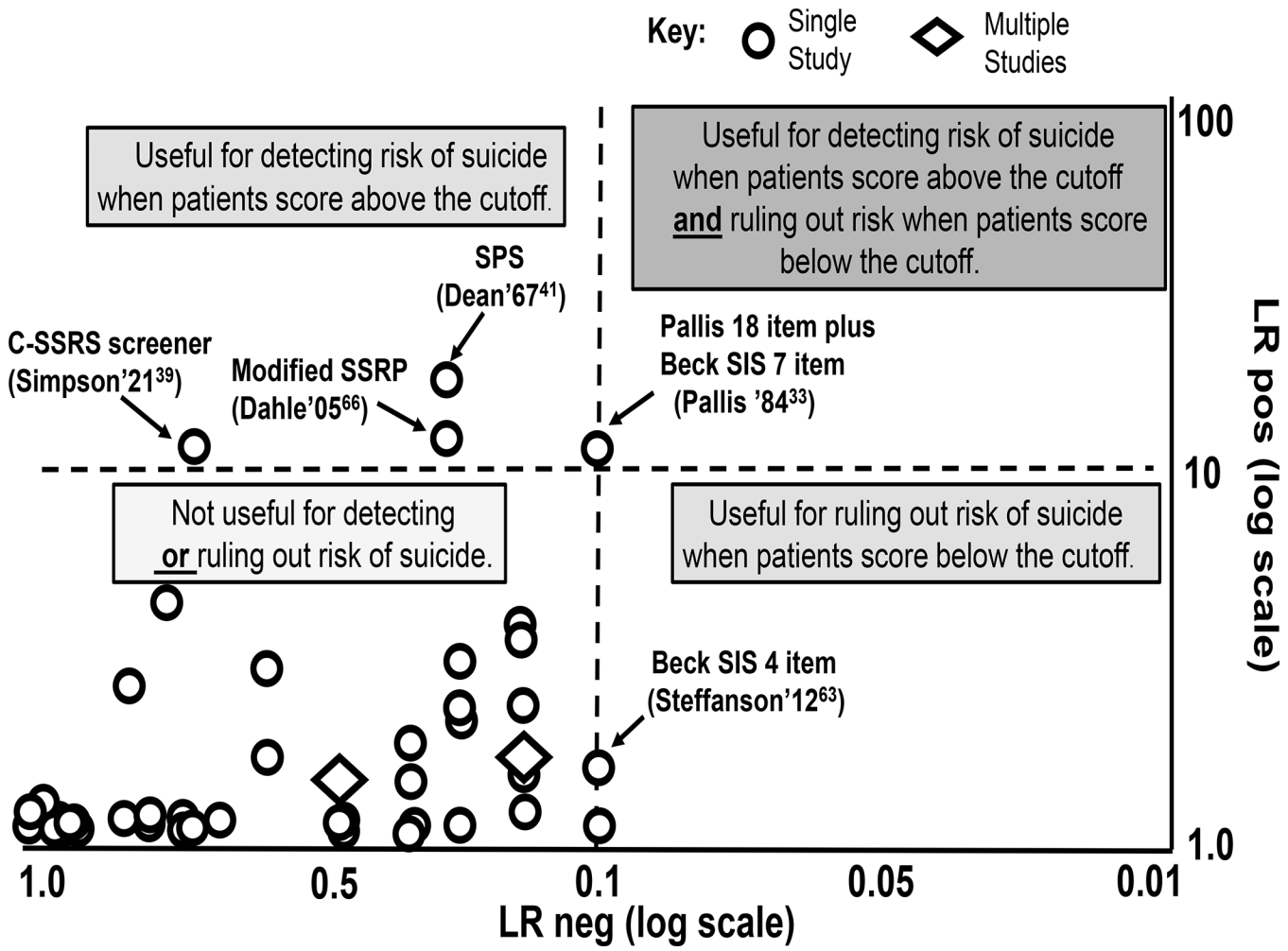

Figure 4:

Likelihood ratio scattergram of instruments to detect or rule out risk of suicide.[a]

[a] To ease the visual interpretation of the results, we have reversed the order of the LR negative values (log scale) such that results are presented from highest to lowest values (i.e., 1.0 – 0.01).

Table 1:

Characteristics of included studies

Author/Year	Country	Setting	Inclusion Criteria	Study Years	Sample N	Suicide N	Follow-up, m	Examined Instrument(s)	Suicide detection
Randomized Controlled Trial									
Ducher'06[48]	France	20 centers; outpatient	≥2 Major depressive episodes; ≥25 MADRS; age 18–70y	NR	103	2	18	RSD	NR
Cohort Studies									
Beck '99[45]	USA	Outpatient mental health	Any mental health condition	1975 – 1994	3,701	30	180	SSI-C; SSI-W	Registry
Buglass'70[34]	United Kingdom	Specialized adult unit	Self-destructive acts	1962–1963, 1967	927	19[a]	36	B & H 3- & 9-item	Registry
Buglass'74[29]	United Kingdom	Specialized adult unit	Parasuicide	1968 – 1970	2,809	23	12	B & H 6-item	Registry
Clark'87[49]	USA	Inpatient mental health	Depression; age 18–70y	NR	593	14	24	Motto RE	Coroner/MD/ Family
Cooper'06[50]	United Kingdom	Emergency Room	Self-harm	1997 – 2001	9,086	22	6	MSHR	Registry[f]
Harris'05[51–52]	United Kingdom	Inpatient mental health	Deliberate self-harm; age ≥ 15y	1993 – 1997	2,489	54	Mean 62.4	SIS	Registry[f]
Katz'17[35]	Canada	Emergency Room[b]	Any mental health condition; adults	2009 – 2013	5,462	77	Median 36	SAD; Modified SAD	Registry[f]
Kuerz' 88[32]	Germany	Inpatient mental health, age ≥ 15y	Deliberate self-harm	1981	421	13	12	GGCS; NSPQ; Pallis; SAD; B & H 6-item	Personal follow-up
Lindh' 19[26]	Sweden	Emergency Room	Self-harm; age 18–95y	2012 – 2016	804	19	12	SIS; C-SSRS; SUAS	Registry[f]
Lopez-Morinigo'18[55]	United Kingdom	Case Register/ Outpatient	Any mental health condition	2007 – 2015	13,758	81	80,769.2 person-year	SLaM	Registry[f]
Motto'85[36–38]	USA	Inpatient mental health	Suicidal state or depression; adults	1969 – 1977	2,753	136	24	Motto RE	Registry/MD/ Coroner/Death certificate/ Family
Naud'10[56]	Canada	Prison	Inmate; adults	1995 –1996; 2006 – 2007	1,047	26	120	S. Probability Scale	Prison authorities/ Coroner
Nimeus'00[57]	Sweden	Inpatient mental health[c,d]	Suicide attempt	1987 – 1997	191	8	12	SUAS	Registry

Author/Year	Country	Setting	Inclusion Criteria	Study Years	Sample N	Suicide N	Follow-up, m	Examined Instrument(s)	Suicide detection
Nimeus '02[59]	Sweden	Inpatient mental health [c]	Suicide attempt	NR	555	22	Mean 54	SIS	Registry
Pallis '84[33]	United Kingdom	Inpatient mental health/Emergency Room	Suicide attempt; three sites with varying age criteria: ≥ 15y, 13–81y; and ≥ 17y	NR	1,263	20	24	Pallis 6 & 18; Modified SIS-7; Pallis 6 + Modified SIS-7; Pallis 18 + Modified SIS-7	Registry [f]
Randall '19[27]	Canada	Emergency Room [b]	Any mental health condition; adults	2009 – 2012	7,872 events	76	12	C-CASA	Registry [f]
Samuelson'06[61]	Sweden	Inpatient mental health	Suicide attempt	NR	15	6	60	SIS	Registry
Simpson '21[39]	USA	Emergency Room	General population; age ≥ 18y	2016 – 2018	92,643 events	63	12	C-SSRS screener	Registry
Steeg '12[28]	United Kingdom	Emergency Room	Self-harm; age ≥ 16y	2003 – 2007	29,571 events	92	6	MSHR; ReACT;	Registry [f]
Steeg '18[62]	United Kingdom	Emergency Room	Self-harm	2010 – 2012	4,000 events	18	6	MSHR; ReACT; SAD; Modified SAD	Registry/ Coroner
Stefansson '12[63]	Sweden	Clinical follow-up after hospitalization	Suicide attempt; age ≥ 18y	1993 – 1998	81	7	Mean 114	SIS; Modified SIS-4	Registry
Wang '85[42]	Denmark	Inpatient mental health	Suicide attempt; age ≥ 15y	1980 – 1984	99	10	41	PIS	Hospital records/death certificate
Case-Control Studies									
Braucht '70[40]	USA	Adult inpatient mental health [e]	Any mental health condition	1961 – 1969	482	63	-	Revised SPS	Hospital records
Dahle '05[66]	Germany	Prison	Pre-trial inmate; adults	1991 – 2000	60	30	-	SSRP; Modified SSRP	Prison records
Dean '67[41]	USA	Adult inpatient mental health [e]	Any mental health condition	1961 – 1965	172	16	-	SPS	Hospital records
Farberow '74[67]	USA	Adult inpatient mental health	U.S. Veterans with any mental health condition	1966 – 1968	379	185	-	NHSPS	Hospital records
Frottier '08[30]	Austria	Prison	Pre-trial and sentenced inmate	1975 – 1999	484	173	-	VISCI	Prison records
Frottier '09[31]	Austria	Prison	Pre-trial and sentenced inmate	2000 – 2004	165	55	-	VISCI	Prison records
Van de Loo '70[19]	Netherlands	Adult inpatient mental health	Suicide attempt	1960 – 1968	209	57	-	NSPQ	Registry/police reports

B & H = Buglass and Horton Scale; C-CASA = Columbia Classification System; C-SSRS screener = Columbia Suicide Severity Rating Scale–Clinical Practice Screener; GGCS = Golden Gate Community Scale; IQR = Interquartile Range; m = months; MADRS = Montgomery-Asberg Depression Rating Scale; MD = Physician; MSHR = Manchester Self-Harm Rule; NHSPS = Neuropsychiatric Hospital Suicide Prediction Schedule; NR = Not reported; NSPQ = Nijmegen Suicide Prediction Questionnaire; PIS = Pierce Intent Scale; RE = Risk Estimator for Suicide; RSD = The suicidal risk assessment scale of Ducher; SAD = SAD PERSONS Scale; SIS = Suicide Intent Scale (Beck), SLaM = South London and Maudsley NHS Foundation Trust; SPS = Suicide Potential Scale; SSI = The Scale of Suicide Ideation (Beck); SSRP = Screening for Suicide Risk of Prisoners; SUAS = Suicide Assessment Scale; S. Probability Scale = Suicide Probability Scale; VISCI = The Viennese Instrument for Suicidality in Correctional Institutions; y = years

[a]Authors reported there were too few suicides in the 1967 cohort to permit meaningful analysis.

[b]Subjects were drawn from same underlying population

[c]Subjects likely came from same underlying population

[d]Study was designed as prospective analysis but controls were selected using case-control methodology.

[e]Subjects were drawn from same underlying population

[f]Definition includes other causes that may be misclassified suicide deaths (e.g., undetermined cause or accidental poisoning or probable suicide).