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

Natalie B. Riblet, MD, MPH[a,b]; Sarah Matsunaga, BA[c]; Younji Lee, BA[c]; Yinong Young-Xu, ScD[a,d]; Brian Shiner, MD, MPH[a,b,e]; Paula P. Schnurr, PhD[d,e]; Maxwell Levis, PhD[a,d]; and Bradley V. Watts, MD, MPH[a,d,f]

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, PsycINFO, 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.

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% CIs 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 department, inpatient mental health, and prison settings when patients scored above the cutoff (LR+ > 10). For example, among patients discharged from an emergency department, the Columbia Suicide Severity Rating Scale–Clinical Practice Screener had a LR+ of 10.3 (95% CI, 6.3–16.8) at 3-month follow-up. The clinical utility of the instruments depends on the pretest probability of suicide in the setting. Because studies spanned over 6 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 department, inpatient mental health, or prison settings. The utility of the instrument hinges, in part, on baseline suicide risk.

Registration: PROSPERO: CRD42021285528

J Clin Psychiatry 2023;84(1):21r14385

To cite: Riblet NB, Matsunaga S, Lee Y. Tools to detect risk of death by suicide: a systematic review and meta-analysis. *J Clin Psychiatry.* 2023;84(1):21r14385.

To share: https://doi.org/10.4088/JCP.21r14385

[a]Mental Health Service, Veterans Affairs Medical Center, White River Junction, Vermont

[b]Department of Psychiatry and Dartmouth Institute, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire

[c]Geisel School of Medicine at Dartmouth, Hanover, New Hampshire

[d]Department of Psychiatry, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire

[e]National Center for PTSD, White River Junction, Vermont

[f]Veterans Rural Health Resource Center, White River Junction VA Medical Center, White River Junction, Vermont

**Corresponding author:* Natalie B. Riblet, MD, MPH, Veterans Affairs Medical Center, Bldg 9, 215 North Main St, White River Junction, VT 05009 (Natalie.Riblet@dartmouth.edu).

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 using a validated instrument in patients evaluated (or treated) for behavioral health conditions.[2] Available instruments generally have high face validity. A series of systematic reviews and meta-analyses,[3–7] however, concluded that there is a lack of robust evidence to support the notion that these instruments can reliably detect risk of death by suicide. 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 (ie, sensitivity > 80% and specificity > 50%).[4] Similarly, Carter et al[7] reported that the pooled positive predictive value (PPV) of psychological instruments to detect risk of suicide was only 5.5%. In a 2013 systematic review completed for the US Preventive Services Task Force, O'Connor et al[8] concluded that there was only minimal evidence to support the routine practice of suicide screening in primary care settings. 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 health care 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 posttest 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 health care provider or health care system can account for

Clinical Points

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

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 on 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 nonfatal 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), PsycINFO, 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 3 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 years 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 (ie, < 10%). If age was reported as a categorical variable (eg, 15–19 years, 20–39 years), 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[19] provided no patient-level characteristics, we contacted the hospital where the study was conducted to confirm that the psychiatric unit was an adult unit (ie, ≥ 18 years) at the time of the study in 1970. 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.

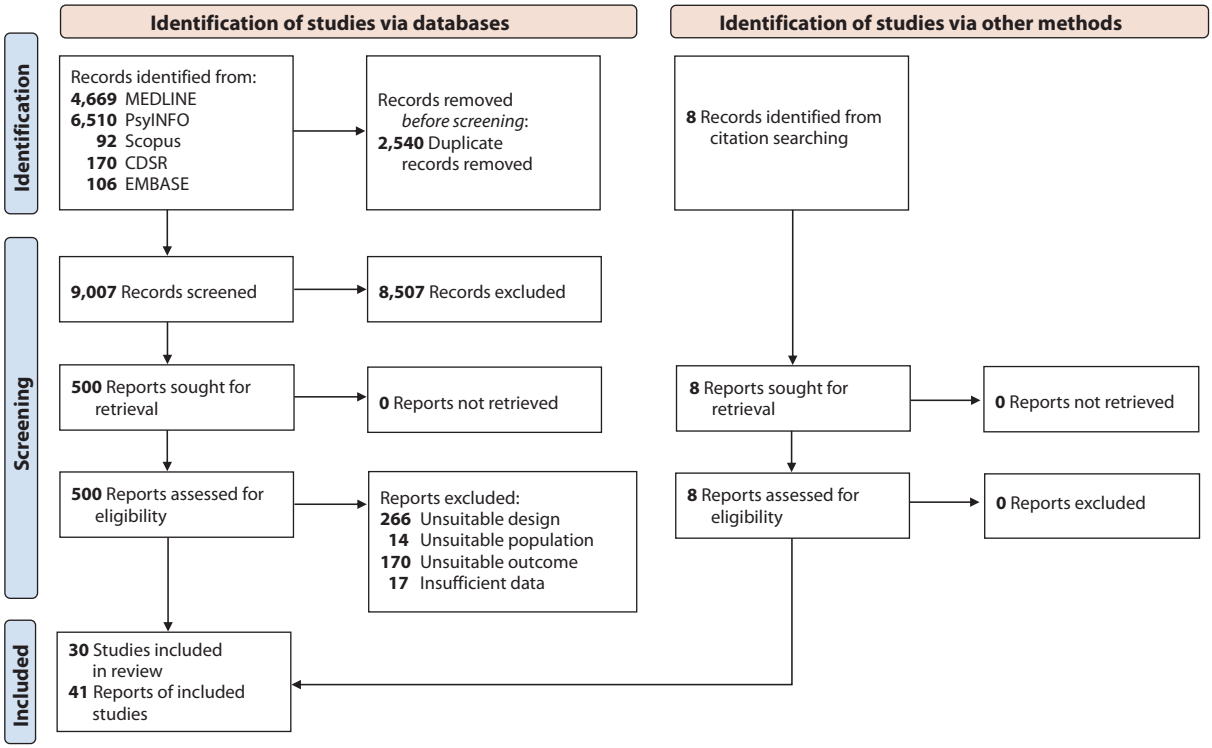
Applying our a priori inclusion criteria, one reviewer (N.B.R.) screened the titles and abstracts of all potentially relevant studies, excluding those that were clearly ineligible. Three reviewers (N.B.R., S.M., Y.L.) then independently and in duplicate assessed the full text of the remaining studies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

Three reviewers (N.B.R., S.M., Y.L.) extracted data in duplicate from included studies. We extracted data related to demographics, methods, outcomes and risk of bias. We used the QUADAS 2 scale[16] to evaluate risk of bias. 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

Figure 1. PRISMA Flow Diagram



Abbreviation: CDSR = Cochrane Database of Systematic Reviews.

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 (CIs).[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 4 or more distinct populations. We applied bivariate mixed effects regression methods to calculate the summary estimates and summary receiver operator curves (SROCs) 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 outcome 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 for 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– for 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– of 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 1 of 4 quadrants to further assess clinical utility: right upper quadrant (exclusion and 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 (Microsoft Corporation).

If a study reported findings by subgroup (eg, 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 authors for these data.[26–28] If we could not obtain the needed data from the authors,[26–28] we excluded the datapoint 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.

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 1 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 shown 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 department population for detecting risk of suicide at 3-month follow-up when patients had a positive screen (ie, "yes" to any of 3 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 department 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 5 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

Table 1. Characteristics of Included Studies

First Author/ Year	Country	Setting	Inclusion Criteria	Study Years	Sample N	Suicide n	Follow-up, mo	Examined Instrument(s)	Suicide Detection
Randomized Controlled Trial									
Ducher 2006[48]	France	20 Centers; outpatient	≥2 major depressive episodes; ≥25 MADRS score; age 18–70 y	NR	103	2	18	RSD	NR
Cohort Studies									
Beck 1999[45]	USA	Outpatient mental health	Any mental health condition	1975–1994	3,701	30	180	SSI-C, SSI-W	Registry
Buglass 1970[34]	United Kingdom	Specialized adult unit	Self-destructive acts	1962–1963, 1967	927	19[a]	36	B & H, 3- and 9-item	Registry
Buglass 1974[29]	United Kingdom	Specialized adult unit	Parasuicide	1968–1970	2,809	23	12	B & H, 6-item	Registry
Clark 1987[49]	USA	Inpatient mental health	Depression; age 18–70 y	NR	593	14	24	Motto RE	Coroner/MD/family
Cooper 2006[50]	United Kingdom	Emergency department	Self-harm	1997–2001	9,086	22	6	MSHR	Registry[b]
Harris 2005[51,52]	United Kingdom	Inpatient mental health	Deliberate self-harm; age ≥15 y	1993–1997	2,489	54	Mean 62.4	SIS	Registry[b]
Katz 2017[35]	Canada	Emergency department[c]	Any mental health condition; adults	2009–2013	5,462	77	Median 36	SAD, Modified SAD	Registry[b]
Kuerz 1988[32]	Germany	Inpatient mental health, age ≥15 y	Deliberate self-harm	1981	421	13	12	GGCS; NSPQ; Pallis; SAD; B & H, 6-item	Personal follow-up
Lindh 2019[26]	Sweden	Emergency department	Self-harm; age 18–95 y	2012–2016	804	19	12	SIS, C-SSRS, SUAS	Registry[b]
Lopez-Morinigo 2018[55]	United Kingdom	Case register/ outpatient	Any mental health condition	2007–2015	13,758	81	80,769.2 Person-years	SLaM	Registry[b]
Motto 1985[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 2010[56]	Canada	Prison	Inmate; adults	1995–1996; 2006–2007	1,047	26	120	Suicide Probability Scale	Prison authorities/ coroner
Niméus 2000[57]	Sweden	Inpatient mental Health[d,e]	Suicide attempt	1987–1997	191	8	12	SUAS	Registry
Niméus 2002[59]	Sweden	Inpatient mental Health[d]	Suicide attempt	NR	555	22	Mean 54	SIS	Registry
Pallis 1984[33]	United Kingdom	Inpatient mental health/emergency department	Suicide attempt; 3 sites with varying age criteria: ≥15 y, 13–81 y; and ≥17 y	NR	1,263	20	24	Pallis, 6- and 18-item; Modified SIS-7; Pallis, 6-item + Modified SIS-7; Pallis, 18-item + Modified SIS-7	Registry[b]
Randall 2019[27]	Canada	Emergency department[c]	Any mental health condition; adults	2009–2012	7,872 events	76	12	C-CASA	Registry[b]
Samuelsson 2006[61]	Sweden	Inpatient mental health	Suicide attempt	NR	15	6	60	SIS	Registry
Simpson 2021[39]	USA	Emergency department	General population; age ≥18 y	2016–2018	92,643 events	63	12	C-SSRS Screener	Registry

(continued)

Table 1 (continued).

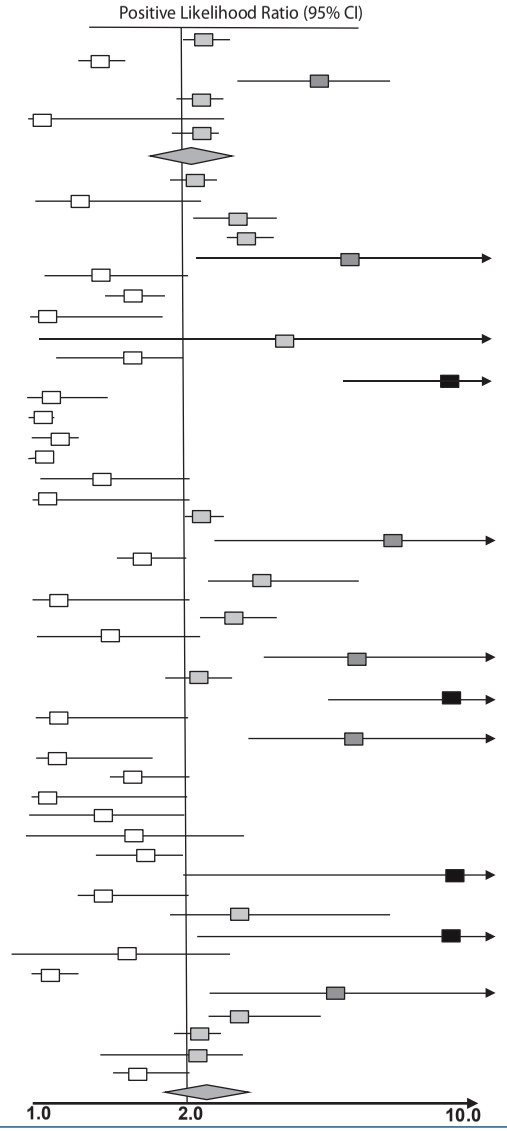
First Author/ Year	Country	Setting	Inclusion Criteria	Study Years	Sample N	Suicide n	Follow-up, mo	Examined Instrument(s)	Suicide Detection
Cohort Studies									
Steeg 2012[28]	United Kingdom	Emergency department	Self-harm; age ≥ 16 y	2003–2007	29,571 events	92	6	MSHR, ReACT	Registry[b]
Steeg 2018[62]	United Kingdom	Emergency department	Self-harm	2010–2012	4,000 events	18	6	MSHR ReACT, SAD, Modified SAD	Registry/coroner
Stefansson 2012[63]	Sweden	Clinical follow-up after hospitalization	Suicide attempt; age ≥ 18 y	1993–1998	81	7	Mean 114	SIS, Modified SIS-4	Registry
Wang 1985[42]	Denmark	Inpatient mental health	Suicide attempt; age ≥ 15 y	1980–1984	99	10	41	PIS	Hospital records/death certificate
Case-Control Studies									
Braucht 1970[40]	USA	Adult inpatient mental health[f]	Any mental health condition	1961–1969	482	63	…	Revised SPS	Hospital records
Dahle 2005[66]	Germany	Prison	Pre-trial inmate; adults	1991–2000	60	30	…	SSRP, Modified SSRP	Prison records
Dean 1967[41]	USA	Adult inpatient mental health[f]	Any mental health condition	1961–1965	172	16	…	SPS	Hospital records
Farberow 1974[67]	USA	Adult inpatient mental health	US Veterans with any mental health condition	1966–1968	379	185	…	NHSPS	Hospital records
Frottier 2008[30]	Austria	Prison	Pre-trial and sentenced inmate	1975–1999	484	173	…	VISCI	Prison records
Frottier 2009[31]	Austria	Prison	Pre-trial and sentenced inmate	2000–2004	165	55	…	VISCI	Prison records
van de Loo 1970[19]	Netherlands	Adult inpatient mental health	Suicide attempt	1960–1968	209	57	…	NSPQ	Registry/police reports

[a]Authors reported there were too few suicides in the 1967 cohort to permit meaningful analysis.
[b]Definition includes other causes that may be misclassified suicide deaths (eg, undetermined cause or accidental poisoning or probable suicide).
[c]Subjects were drawn from same underlying population.
[d]Subjects likely came from same underlying population.
[e]Study was designed as prospective analysis but controls were selected using case-control methodology.
[f]Subjects were drawn from same underlying population.
Abbreviations: 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, 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 = 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 = Scale of Suicide Ideation (Beck), SSRP = Screening for Suicide Risk of Prisoners, SUAS = Suicide Assessment Scale, VISCI = Viennese Instrument for Suicidality in Correctional Institutions.

Figure 2. Ability of Instruments to Detect Risk of Suicide When Patients Score at or Above the Cutoff

Name of Scale	First Author/Year	Scale Items: SI/SB	Clin-Dem	MH Symptom	Follow-up, mo	Cutoff	SN/SP	LR+ (95% CI)
Beck SIS Total	Harris 2005 (female)[51,52]	X	-	-	Mean 62.4	14+	67/75	2.7 (2.0–3.6)
	Harris 2005 (male)[51,52]	X	-	-	Mean 62.4	10+	77/49	1.5 (1.2–1.8)
	Lindh 2019[26]	X	-	-	3	21+	100/82	4.9 (3.4–7.0)
	Niméus 2002[59]	X	-	-	Mean 54	19+	59/77	2.6 (1.8–3.8)
	Samuelsson 2006[61]	X	-	-	12	19+	40/50	0.8 (0.2–2.8)
	Stefansson 2012[63]	X	-	-	Mean 114	16+	100/52	1.9 (1.4–2.6)
					Summary Estimate		**69/67**	**2.1 (1.6–2.6)**
Beck SIS, 4-item	Stefansson 2012[63]	X	-	-	Mean 114	6+	100/59	2.3 (1.6–3.2)
Beck SIS, 7-item	Pallis 1984[33]	X	-	-	6	4+	50/60	1.3 (0.6–2.8)
Beck SSI, current	Beck 1999[45]	X	-	-	180	2+	53/83	3.1 (2.2–4.3)
Beck SSI, worst	Beck 1999[45]	X	-	-	180	16+	80/78	3.6 (3.0–4.3)
B & H, 3-item	Buglass 1970 (male)[34]	X	X	-	12	3+	50/91	5.3 (2.1–13.9)
B & H, 6-item	Buglass 1974 (index)[29]	X	X	-	12	1+	86/44	1.5 (1.1–2.1)
	Buglass 1974 (validation)[29]	X	X	-	12	1+	97/41	1.6 (1.5–1.8)
	Kuerz 1988[32]	X	X	-	12	1+	54/50	1.1 (0.6–1.8)
B & H, 9-item	Buglass 1970 (female)[34]	X	X	-	36	4+	29/93	4.2 (1.2–14.6)
C-CASA	Randall 2019[27]	X	-	-	12	+3-item	70/55	1.6 (1.2–2.0)
C-SSRS Screener	Simpson 2021[39]	X	-	-	3	+Scn	37/96	10.3 (6.3–16.8)
Decision Rule—MSHR	Cooper 2006 (validation)[50,a]	X	X	-	6	1+	100/26	1.2 (0.9–1.7)
	Steeg 2018[62]	X	X	-	6	1+	89/11	1.0 (0.99–1.0)
Decision Rule—ReACT	Steeg 2012 (validation)[28,a]	X	X	-	6	1+	88/24	1.2 (1.0–1.3
	Steeg 2018[62]	X	X	-	6	1+	78/18	1.0 (0.9–1.0)
GGCS	Kuerz 1988[32]	X	X	-	12	5+	54/61	1.4 (0.8–2.3)
Motto RE	Clark 1987[49]	X	X	X	24	8th D+	21/74	0.8 (0.3–2.2)
	Motto 1985[36–38]	X	X	X	24	4.9%	71/72	2.5 (2.2–2.8)
NHSPS	Farberow 1974 (validation)[67,a]	X	X	X	Case-control	130+	38/95	7.6 (3.1–18.6)
NSPQ	Kuerz 1988[32]	X	X		12	4+	92/47	1.7 (1.5–2.1)
	van de Loo 1970[19]	X	X		Case-control	5+	77/79	3.7 (2.6–5.1)
Pallis, 6-item	Kuerz 1988[32]	X	X	-	12	-0.5	46/61	1.2 (0.7–2.2)
	Pallis 1984[33]	X	X	-	12	27.5	83/74	3.2 (2.4–4.2)
Pallis, 18-item	Kuerz 1988[32]	X	X	X	12	0.25	54/62	1.4 (0.8–2.4)
	Pallis 1984[33]	X	X	X	6	86.5	100/86	6.3 (3.5–11.5)
Pallis, 6-item + Beck SIS, 7-item	Pallis 1984[33]	X	X	-	24	e+	72/69	2.3 (1.7–3.2
Pallis, 18-item + Beck SIS, 7-item	Pallis 1984[33]	X	X	X	12	g+	100/91	10.1 (5.4–18.6
PIS	Wang 1985[42]	X	-	-	41	10+	60/49	1.2 (0.7–2.1)
RSD	Ducher 2006[48]	X	-	-	18	7+	100/87	6.2 (2.8–13.8)
SAD PERSONS Scale	Katz 2017[35]	X	X	X	12	5+	49/60	1.2 (0.9–1.7)
	Kuerz 1988[32]	X	X	X	12	3+	69/57	1.6 (1.1–2.4)
	Steeg 2018[62]	X	X	X	6	5+	33/70	1.1 (0.6–2.1)
Modified SAD PERSONS Scale	Katz 2017[35]	X	X	X	6	6+	58/60	1.4 (1.0–2.0)
	Steeg 2018[62]	X	X	X	6	6+	17/90	1.6 (0.6–4.5)
SLaM	Lopez-Morinigo 2018[55]	X	X	X	96	4–5+	65/62	1.7 (1.5–2.0)
SPS	Dean 1967[41]	X	X	X	Case-control	33+	75/94	12 (1.8–81.7)
Revised SPS	Braucht 1970[40]	X	X	X	Case-control	NR	44/68	1.4 (0.9–2.1
SSRP	Dahle 2005[66]	X	X	-	Case-control	24+	83/77	3.6 (1.8–7.0)
Modified SSRP	Dahle 2005[66]	X	X	-	Case-control	3+	70/93	10.5 (2.7–40.9)
Suicide Probability Scale	Naud 2010[56]	X	X		120	50+	23/85	1.5 (0.7–3.1)
SUAS	Lindh 2019[26]	X	-	X	12	39+	61/31	0.9 (0.6–1.3)
	Niméus 2000[57]	X	-	X	12	39+	75/86	5.0 (2.2–11.6)
VISCI	Frottier 2008 (pre-trial)[30]	X	X	-	Case-control	3.12	80/78	3.7 (2.6–5.3)
	Frottier 2008 (sentenced)[30]	X	X	-	Case-control	1.08	89/62	2.3 (1.9–2.8
	Frottier 2009 (pre-trial)[31]	X	X	-	Case-control	2.52	74/65	2.1 (1.4–3.1)
	Frottier 2009 (sentenced)[31]	X	X	-	Case-control	1.08	88/45	1.6 (1.2–2.2)
					Summary Estimate		**85/65**	**2.4 (1.8–3.2)**

Key: No utility; Small-moderate utility; High utility



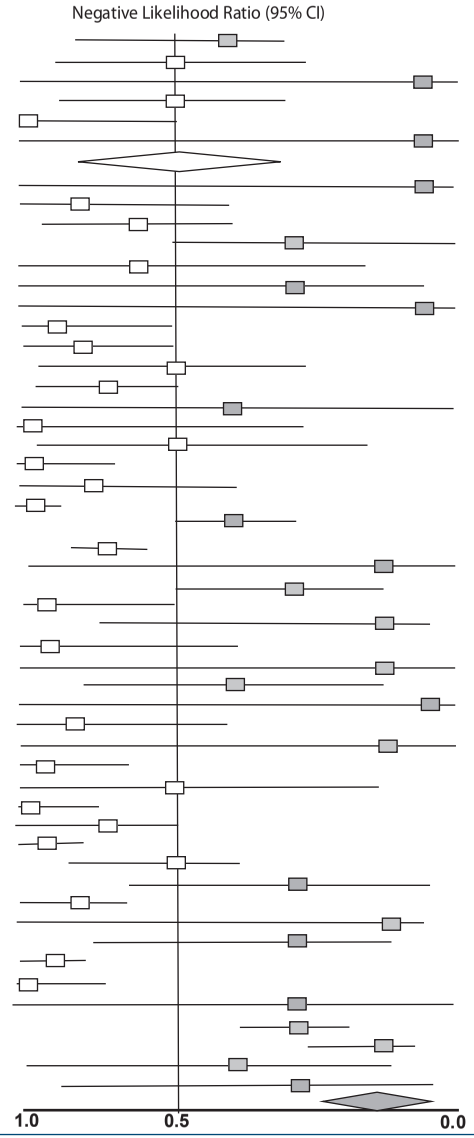
[a]The studies found that LR+ were similar across derivation and validation sets.

Abbreviations: B & H = Buglass and Horton Scale, C-CASA = Columbia Classification System, Clin-Dem = clinical and/or demographic characteristics, C-SSRS Screener = Columbia Suicide Severity Rating Scale–Clinical Practice Screener, D = decile, GGCS = Golden Gate Community Scale, LR = likelihood ratio, MH = mental health, MSHR = Manchester Self-Harm Rule, NHSPS = Neuropsychiatric Hospital Suicide Prediction Schedule, NSPQ = Nijmegen Suicide Prediction Questionnaire, PIS = Pierce Intent Scale, RE = Risk Estimator for Suicide, ReACT = ReACT Self-Harm Rule, RSD = Suicidal Risk Assessment Scale of Ducher, Scn = Screen, SIS = Suicide Intent Scale, SI/SB = suicidal ideation and/or suicidal behavior, SLaM = South London and Maudsley NHS Foundation Trust, SN = sensitivity, SP = specificity, SPS = Suicide Potential Scale, SSI = Scale of Suicide Ideation, SSRP = Screening for Suicide Risk of Prisoners, SUAS = Suicide Assessment Scale, VISCI = Viennese Instrument for Suicidality in Correctional Institutions.

Figure 3. Ability of Instruments to Rule out Risk of Suicide When Patients Score Below the Cutoff

Name of Scale	First Author/Year	Scale Items: SI/SB	Clin-Dem	MH Symptom	Follow-up, mo	Cutoff	SN/SP	LR– (95% CI)
Beck SIS, Total	Harris 2005 (female)[51,52]	X	-	-	Mean 62.4	14+	67/75	0.4 (0.3–0.8)
	Harris 2005 (male)[51,52]	X	-	-	Mean 62.4	10+	77/49	0.5 (0.2–0.9)
	Lindh 2019[26]	X	-	-	3	21+	100/82	0.1 (0.0–1.9)
	Niméus 2002[59]	X	-	-	Mean 54	19+	59/77	0.5 (0.3–0.9)
	Samuelsson 2006[61]	X	-	-	12	19+	40/50	1.2 (0.5–3.1)
	Steffansson 2012[63]	X	-	-	Mean 114	16+	100/52	0.1 (0.0–1.9)
						Summary Estimate	**69/67**	**0.5 (0.3–0.7)**
Beck SIS, 4-item	Stefansson 2012[63]	X	-	-	Mean 114	6+	100/59	0.1 (0.0–1.7)
Beck SIS, 7-item	Pallis 1984[33]	X	-	-	6	4+	50/60	0.8 (0.4–1.9)
Beck SSI, current	Beck 1999[45]	X	-	-	180	2+	53/83	0.6 (0.4–0.8)
Beck SSI, worst	Beck 1999[45]	X	-	-	180	16+	80/78	0.3 (0.1–0.5)
B & H, 3 item	Buglass 1970 (male)[34]	X	X	-	12	3+	50/91	0.6 (0.2–1.2)
B & H, 6 item	Buglass 1974 (index)[29]	X	X	-	12	1+	86/44	0.3 (0.1–2.0)
	Buglass 1974 (validation)[29]	X	X	-	12	1+	97/41	0.1 (0.0–1.2)
	Kuerz 1988[32]	X	X	-	12	1+	54/50	0.9 (0.5–1.7)
B & H, 9 item	Buglass 1970 (female)[34]	X	X	-	36	4+	29/93	0.8 (0.5–1.2)
C-CASA	Randall 2019[27]	X	-	-	12	+3 item	70/55	0.5 (0.3-0.9)
C-SSRS Screener	Simpson 2021[39]	X	-	-	3	+Scn	37/96	0.7 (0.5–0.9)
Decision Rule—MSHR	Cooper 2006 (validation)[50,a]	X	X	-	6	1+	100/26	0.4 (0.03–5.8)
	Steeg 2018[62]	X	X	-	6	1+	89/11	1.0 (0.4–3.0)
Decision Rule—ReACT	Steeg 2012 (validation)[28,a]	X	X	-	6	1+	88/24	0.5 (0.2–1.4)
	Steeg 2018[62]	X	X	-	6	1+	78/18	1.2 (0.7–2.0)
GGCI	Kuerz 1988[32]	X	X	-	12	5+	54/61	0.8 (0.4–1.4)
Motto RE	Clark 1987[49]	X	X	X	24	8th D+	21/74	1.1 (0.8–1.4)
	Motto 1985[36–38]	X	X	X	24	4.9%	71/72	0.4 (0.3–0.5)
NHSPS	Farberow 1974 (validation)[67,a]	X	X	X	Case-control	130+	38/95	0.7 (0.6–0.8)
NSPQ	Kuerz 1988[32]	X	X		12	4+	92/47	0.2 (0.0–1.1)
	van de Loo 1970[19]	X	X		Case-control	5+	77/79	0.3 (0.2-0.5)
Pallis, 6-item	Kuerz 1988[32]	X	X	-	12	-0.5	46/61	0.9 (0.5–1.5)
	Pallis 1984[33]	X	X	-	12	27.5	83/74	0.2 (0.1–0.8)
Pallis, 18-item	Kuerz 1988[32]	X	X	X	12	0.25	54/62	0.7 (0.4–1.3)
	Pallis 1984[33]	X	X	X	6	86.5	100/86	0.2 (0.0–2.2)
Pallis, 6-item + Beck SIS, 7-item	Pallis 1984[33]	X	X	-	24	e+	72/69	0.4 (0.2–0.8)
Pallis, 18-item + Beck SIS, 7-item	Pallis 1984[33]	X	X	X	12	g+	100/91	0.1 (0.0–1.7)
PIS	Wang 1985[42]	X	-	-	41	10+	60/49	0.8 (0.4–1.8)
RSD	Ducher 2006[48]	X	-	-	18	7+	100/87	0.2 (0.0–2.7)
SAD PERSONS Scale	Katz 2017[35]	X	X	X	12	5+	49/60	0.9 (0.6–1.2)
	Kuerz 1988[32]	X	X	X	12	3+	69/57	0.5 (0.2–1.2)
	Steeg 2018[62]	X	X	X	6	5+	33/70	1.0 (0.7–1.3)
Modified SAD PERSONS Scale	Katz 2017[35]	X	X	X	6	6+	58/60	0.7 (0.5–1.1)
	Steeg 2018[62]	X	X	X	6	6+	17/90	0.9 (0.8–1.1)
SLAM	Lopez-Morinigo 2018[55]	X	X	X	96	4–5+	65/62	0.5 (0.4–0.7)
SPS	Dean 1967[41]	X	X	X	Case-control	33+	75/94	0.3 (0.1-0.6)
Revised SPS	Braucht 1970[40]	X	X	X	Case-control	NR	44/68	0.8 (0.6–1.1)
SSRP	Dahle 2005[66]	X	X	-	Case-control	24+	83/77	0.2 (0.1–0.5)
Modified SSRP	Dahle 2005[66]	X	X	-	Case-control	3+	70/93	0.3 (0.2–0.6)
Suicide Probability Scale	Naud 2010[56]	X		X	120	50+	23/85	0.9 (0.7–1.1)
SUAS	Lindh 2019[26]	X	-	X	12	39+	61/31	1.3 (0.7–2.3)
	Niméus 2000[57]	X	-	X	12	39+	75/86	0.3 (0.1–1.0)
VISCI	Frottier 2008 (pre-trial)[30]	X	X	-	Case-control	3.12	80/78	0.3 (0.2–0.4)
	Frottier 2008 (sentenced)[30]	X	X	-	Case-control	1.08	89/62	0.2 (0.1–0.3)
	Frottier 2009 (pre-trial)[31]	X	X	-	Case-control	2.52	74/65	0.4 (0.2–0.8)
	Frottier 2009 (sentenced)[31]	X	X	-	Case-control	1.08	88/45	0.3 (0.1–0.8)
						Summary Estimate	**85/65**	**0.2 (0.17–0.3)**

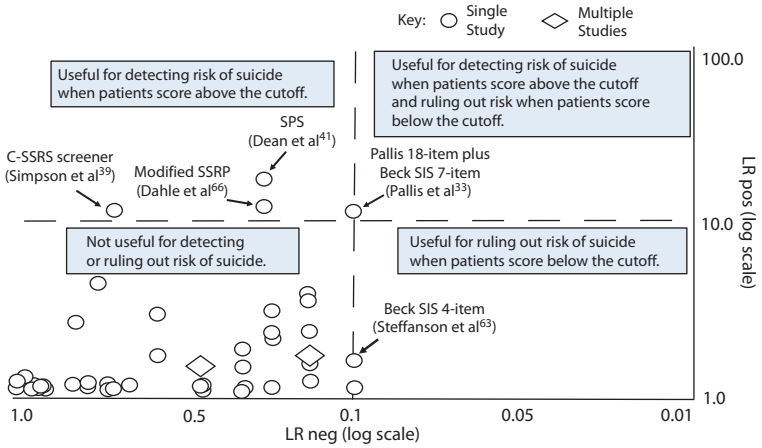
Key: No utility; Small-moderate utility; High utility



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

Abbreviations: B & H = Buglass and Horton Scale, C-CASA = Columbia Classification System, Clin-Dem = clinical and/or demographic characteristics, C-SSRS Screener = Columbia Suicide Severity Rating Scale–Clinical Practice Screener, D = decile, GGCS = Golden Gate Community Scale, LR = likelihood ratio, MH = mental health, MSHR = Manchester Self-Harm Rule, NHSPS = Neuropsychiatric Hospital Suicide Prediction Schedule, NSPQ = Nijmegen Suicide Prediction Questionnaire, PIS = Pierce Intent Scale, RE = Risk Estimator for Suicide, ReACT = ReACT Self-Harm Rule, RSD = Suicidal Risk Assessment Scale of Ducher, Scn = Screen, SIS = Suicide Intent Scale, SI/SB = suicidal ideation and/or suicidal behavior, SLaM = South London and Maudsley NHS Foundation Trust, SN = sensitivity, SP = specificity, SPS = Suicide Potential Scale, SSI = Scale of Suicide Ideation, SSRP = Screening for Suicide Risk of Prisoners, SUAS = Suicide Assessment Scale, VISCI = Viennese Instrument for Suicidality in Correctional Institutions.

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 (ie, 1.0–0.01).
Abbreviations: C-SSRS Screener = Columbia Suicide Severity Rating Scale–Clinical Practice Screener, LR = likelihood ratio, neg = negative, pos = positive, SIS = Suicide Intent Scale, SPS = Suicide Potential Scale, SSI = Scale of Suicide Ideation, SSRP = Screening for Suicide Risk of Prisoners.

(sentenced and pretrial), 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 (eg, LR+, 13) and index sample (eg, 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 1- (LR+, 5.0; LR–, 0.8), 6- (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[37] reported an extremely high LR+ > 50 for the Motto Risk Estimate for detecting risk of suicide at 12-month follow-up among hospitalized patients. The authors were unable to replicate this result in a validation sample at 12 months 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 (ie, Beck SIS 4-item, Beck SSI current and worst, C-CASA, MSHR, ReACT, and Pallis 6-item item + Beck SIS 7-item) or moderate (ie, Beck SIS 7-item, Buglass and Horton 3- and 6-item, C-SSRS Screener, Neuropsychiatric Hospital Suicide Prediction Schedule, Pallis 18-item item + Beck SIS 7-item, PIS, 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 department 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[39] examined the C-SSRS Screener in 92,643 patients who presented to an emergency room with any acute health concern. While the authors identified that a positive screen had a high utility for detecting risk of suicide at 90-day follow-up (ie, LR+, 10.3), the instrument performed poorly

at ruling out cases (ie, 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 posttest probability of suicide (ie, 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[68] reported that the probability of suicide within the first 3 months of emergency department discharge among patients who presented with suicidal behavior was greater than 0.65%. This means that a LR+ of 10.3 would increase the posttest 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 be useful only in a population of patients whose baseline suicide risk is high (eg, 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[66] 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). 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 (ie, 70). Frottier et al[30,31] 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+ values ranging from 13 to 38. 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 in the general population,[73] this finding may be clinically meaningful. For example, if the pretest probability of suicide in the first 3 months after psychiatric discharge is 1.1%,[73] then the instrument would increase the posttest 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 (ie, 0.1%.). While this result is quite promising, several factors must be considered. By design, the Beck SIS can be administered only 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 department 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[33] conducted their study more than 35 years ago, and the rates of suicide in the population have shifted over this timeframe.[74–76] 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[4] also concluded that these scales had low diagnostic accuracy to detect risk of suicide. Notably, Runeson et al[4] recommended that the SAD PERSONS scale and its modified version should not be used in their current format. 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 department, Steeg et al[62] determined that the sensitivity of the MHSR was 89, while the specificity was only 11.

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 6

decades in several countries, it is possible that secular trends may have influenced our results. 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, their doing so 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 health care professionals such as emergency department personnel may have little or no contact with a patient during any given year. Therefore, an emergency department clinician would benefit more from knowing about the patient's near-term risk in the next 30 days or 60 days versus the next 2 or 3 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 review 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 2 cases in which 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 (eg, 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 department 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 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 (eg, primary care), but also to develop new and better ways for providers to detect risk of suicide in patients in real-time.

Submitted: January 13, 2022; accepted June 29, 2022.

Published online: November 16, 2022.

Relevant financial relationships: None.

Funding/support: Dr Riblet has received 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).

Role of the sponsor: The supporters had no role in the design, analysis, interpretation, or publication of this study.

Previous presentation: None.

Disclaimer: The views expressed in this article do not necessarily represent the views of the Department of Veterans Affairs or of the United States Government.

Supplementary material: Available at Psychiatrist.com.

REFERENCES

1. Brown G. A review of suicide assessment measures for intervention research with adults and older adults. SPRC website. 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. The Joint Commissiion website. 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. *PLoS One*. 2016;11(6):e0156322.
4. Runeson B, Odeberg J, Pettersson A, et al. *PLoS One*. 2017;12(7):e0180292.
5. Saab MM, Murphy M, Meehan E, et al. *Arch Suicide Res*. 2021:1–21.
6. Chan MK, Bhatti H, Meader N, et al. *Br J Psychiatry*. 2016;209(4):277–283.
7. Carter G, Milner A, McGill K, et al. *Br J Psychiatry*. 2017;210(6):387–395.
8. O'Connor E, Gaynes BN, Burda BU, et al. *Ann Intern Med*. 2013;158(10):741–754.
9. Johnson KM. *Diagnosis (Berl)*. 2017;4(3):159–167.
10. Johnson KM. *Diagnosis (Berl)*. 2018;5(2):89.
11. Deeks JJ, Altman DG. *BMJ*. 2004;329(7458):168–169.
12. Aitken C, Mavridis D. *Evid Based Ment Health*. 2019;22(1):44–48.
13. Cai W, Zhu D, Byanju S, et al. *Medicine (Baltimore)*. 2019;98(14):e14891.
14. Downie A, Williams CM, Henschke N, et al. *BMJ*. 2013;347(dec11 1):f7095.
15. Glenn CR, Nock MK. *Am J Prev Med*. 2014;47(3 suppl 2):S176–S180.
16. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. *Ann Intern Med*. 2011;155(8):529–536.
17. McInnes MDF, Moher D, Thombs BD, et al. The PRISMA-DTA Group. *JAMA*. 2018;319(4):388–396.
18. Macaskill P, Gatsonis C, Deeks J, et al. Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0*, Chapter 10. The Cochrane Collaboration. https://methods.cochrane.org/sdt/handbook-dta-reviews. 2010. Accessed December 9, 2021.
19. van de Loo KJ, Diekstra RW. *Ned Tijdschr Psychol*. 1970;25(2):95–100.
20. Ouzzani M, Hammady H, Fedorowicz Z, et al. *Syst Rev*. 2016;5(1):210.
21. Simel DL, Samsa GP, Matchar DB. *J Clin Epidemiol*. 1991;44(8):763–770.
22. Stengel D, Bauwens K, Sehouli J, et al. *J Med Screen*. 2003;10(1):47–51.
23. Dwamena B, Sylvester R, Carlos RC. MIDAS: meta-analysis of diagnostic accuracy studies. BC.EDU website. http://fmwww.bc.edu/repec/bocode/m/midas.pdf. Accessed October 21, 2021.
24. Reitsma JB, Glas AS, Rutjes AWS, et al. *J Clin Epidemiol*. 2005;58(10):982–990.
25. Jaeschke R, Guyatt GH, Sackett DL. *JAMA*. 1994;271(9):703–707.
26. Lindh ÅU, Dahlin M, Beckman K, et al. *J Clin Psychiatry*. 2019;80(6):18m12707.
27. Randall JR, Sareen J, Chateau D, et al. *Suicide Life Threat Behav*. 2019;49(4):941–951.
28. Steeg S, Kapur N, Webb R, et al. *Psychol Med*. 2012;42(11):2383–2394.
29. Buglass D, Horton J. *Br J Psychiatry*. 1974;124(0):573–578.
30. Frottier P, König F, Matschnig T, et al. (Das Wiener Instrument für Suizidgefahr in Haft). *Psychiatr Prax*. 2008;35(1):21–27.
31. Frottier P, Koenig F, Seyringer M, et al. *Suicide Life Threat Behav*. 2009;39(4):376–385.
32. Kuerz A, Moeller HJ, Torhorst A, et al. *Validation of Six Risk Scales for Suicide Attempters*. Current

Issues of Suicidology; 1988:174–178.
33. Pallis DJ, Gibbons JS, Pierce DW. *Br J Psychiatry*. 1984;144(2):139–148.
34. Buglass D, McCulloch JW. *Br J Psychiatry*. 1970;116(534):483–491.
35. Katz C, Randall JR, Sareen J, et al. *Depress Anxiety*. 2017;34(9):809–816.
36. Motto JA. *Suicide Life Threat Behav*. 1985;15(3):139–150.
37. Motto JA, Heilbron DC. *Development and Validation of Scales for Estimation of Suicidal Risk*. Suicidology; 1976:166–199.
38. Motto JA, Heilbron DC, Juster RP. *Am J Psychiatry*. 1985;142(6):680–686.
39. Simpson S, Goans C, Loh R, et al. *Acad Emerg Med*. 2021;28(6):621–629.
40. Braucht GN, Wilson LT. *J Consult Clin Psychol*. 1970;35(3):426.
41. Dean RA, De Cook R, Maley RF, et al. *J Clin Psychol*. 1967;23(3):296–301.
42. Wang AG, Nielsen B, Bille-Brahe U, et al. *Acta Psychiatr Scand*. 1985;72(4):389–394.
43. GRADEpro GDT: GRADEpro Guideline Development Tool [Software]. McMaster University and Evidence Prime. Gradepro website. https://www.gradepro.org. Accessed December 14, 2021.
44. Schünemann H, Brożek J, Guyatt G, et al, eds. *GRADE Handbook for Grading Quality of Evidence and Strength of Recommendations. Updated October 2013*. The GRADE Working Group. guidelinedevelopment.org/handbook. 2013. Accessed December 14, 2021.
45. Beck AT, Brown GK, Steer RA, et al. *Suicide Life Threat Behav*. 1999;29(1):1–9.
46. Brown GK, Beck AT, Steer RA, et al. *J Consult Clin Psychol*. 2000;68(3):371–377.
47. Joiner TE Jr, Steer RA, Brown G, et al. *Behav Res Ther*. 2003;41(12):1469–1480.
48. Ducher JL, Terra JL. (L'échelle d'évaluation du risque suicidaire RSD possède-t-elle une valeur predictive?) *LEncphale*. 2006;32(5Pt1):738–745.
49. Clark DC, Young MA, Scheftner WA, et al. *Am J Psychiatry*. 1987;144(7):923–926.
50. Cooper J, Kapur N, Dunning J, et al. *Ann Emerg Med*. 2006;48(4):459–466.
51. Harriss L, Hawton K. *J Affect Disord*. 2005;86(2–3):225–233.
52. Harriss L, Hawton K, Zahl D. *Br J Psychiatry*. 2005;186(1):60–66.
53. Lindh AU, Waern M, Beckman K, et al. *BMC Psychiatry*. 2018;18(1):319.
54. Lindh ÅU, Beckman K, Carlborg A, et al. *J Affect Disord*. 2020;263:445–449.
55. Lopez-Morinigo JD, Fernandes AC, Shetty H, et al. *Soc Psychiatry Psychiatr Epidemiol*. 2018;53(11):1161–1171.
56. Naud H, Daigle MS. *J Psychopathol Behav Assess*. 2010;32(3):333–342.
57. Niméus A, Alsén M, Träskman-Bendz L. *Eur Psychiatry*. 2000;15(7):416–423.
58. Skogman K, Alsén M, Ojehagen A. *Soc Psychiatry Psychiatr Epidemiol*. 2004;39(2):113–120.
59. Niméus A, Alsén M, Träskman-Bendz L. *Arch Suicide Res*. 2002;6(3):211–219.
60. Probert-Lindström S, Berge J, Westrin Å, et al. *BMJ Open*. 2020;10(10):e038794.
61. Samuelsson M, Jokinen J, Nordström AL, et al. *Acta Psychiatr Scand*. 2006;113(1):44–47.
62. Steeg S, Quinlivan L, Nowland R, et al. *BMC Psychiatry*. 2018;18(1):113.
63. Stefansson J, Nordström P, Jokinen J. *J Affect Disord*. 2012;136(1–2):167–171.
64. Stefansson J, Nordstrom P, Jokinen J. . *Eur Psychiatry*. 2011;26(S2):1649.
65. Stefansson J, Nordström P, Runeson B, et al. *BMC Psychiatry*. 2015;15(1):226.
66. Dahle K-P, Lohner JC, Konrad N. *Int J Forensic Ment*. 2005;4(1):53–62.
67. Farberow NL, MacKinnon D. *J Nerv Ment Dis*. 1974;158(6):408–419.
68. Geulayov G, Casey D, Bale L, et al. *Lancet Psychiatry*. 2019;6(12):1021–1030.
69. Brådvik L. *Int J Environ Res Public Health*. 2018;15(9):2028.
70. Kajeepeta S, Mauro PM, Keyes KM, et al. *Lancet Public Health*. 2021;6(4):e240–e248.
71. Bukten A, Stavseth MR. *Eur J Epidemiol*. 2021;36(10):1075–1083.
72. Zhong S, Senior M, Yu R, et al. *Lancet Public Health*. 2021;6(3):e164–e174.
73. Chung DT, Ryan CJ, Hadzi-Pavlovic D, et al. *JAMA Psychiatry*. 2017;74(7):694–702.
74. Hedegaard H, Curtin SC, Warner M. *NCHS Data Brief*. 2020;362:1–8.
75. Naghavi M; Global Burden of Disease Self-Harm Collaborators. *BMJ*. 2019;364:l94.
76. McKeown RE, Cuffe SP, Schulz RM. *Am J Public Health*. 2006;96(10):1744–1751.
77. Mann JJ, Michel CA, Auerbach RP. *Am J Psychiatry*. 2021;178(7):611–624.
78. Riblet NBV, Shiner B, Young-Xu Y, et al. *Br J Psychiatry*. 2017;210(6):396–402.
79. Haroz EE, Kitchen C, Nestadt PS, et al. *Suicide Life Threat Behav*. 2021;51(6):1189–1202.
80. Snowdon J, Choi NG. *Glob Public Health*. 2020;15(12):1894–1901.

See supplementary material for this article at PSYCHIATRIST.COM.

THE JOURNAL OF
CLINICAL PSYCHIATRY
THE OFFICIAL JOURNAL OF THE AMERICAN SOCIETY OF CLINICAL PSYCHOPHARMACOLOGY

Supplementary Material

Article Title: Tools to Detect Risk of Death by Suicide: A Systematic Review and Meta-Analysis

Author(s): Natalie B. Riblet, MD, MPH; Sarah Matsunaga, BA; Younji Lee, BA; Yinong Young-Xu, ScD; Brian Shiner, MD, MPH; Paula P. Schnurr, PhD; Maxwell Levis, PhD; and Bradley V. Watts, MD, MPH

DOI Number: https://doi.org/10.4088/JCP.21r14385

List of Supplementary Material for the article

1. Table 1 Quality assessment of included studies

Disclaimer

This Supplementary Material has been provided by the author(s) as an enhancement to the published article. It has been approved by peer review; however, it has undergone neither editing nor formatting by in-house editorial staff. The material is presented in the manner supplied by the author.

Supplementary Table 1: Quality assessment of included studies[a]

	Attrition	Risk of Bias				Applicability Concerns		
Author/Year	**%**	**Selection**	**Index Test**	**Reference Std**	**Flow/Timing**	**Selection**	**Index Test**	**Reference Std**
Randomized Controlled Studies								
Ducher'06[48]	0%	+	–	–	+	+	+	+
Cohort Studies								
Beck'99[45]	0%	+	+	+	+	+	+	+
Buglass'70[34]	0%[b]	+	–	+	+	+	+	+
Buglass'74[29]	0%[b]	+	–	+	+	+	+	+
Clark '87[49]	4%	+	–	–	+	+	+	+
Cooper'06[50]	0%	+	–	+	+	+	+	+
Harris'05[51-52]	8.5%	–	–	+	+	+	+	+
Katz'17[35]	4%	+	–	+	+	+	+	+
Kuerz'88[32]	18%	+	–	X	X	+	+	+
Lindh'19[26]	0%	+	–	+	+	+	+	+
Lopez-Morinigo'18[55]	0%	X	–	+	+	+	+	+
Motto'85[36-38]	0%	+	+	+	+	+	+	+
Naud'10[56]	2%	+	+	+	+	+	+	+
Nimeus'00[57]	0%[c]	X	–	+	+	+	+	+
Nimeus'02[59]	18%[d]	+	–	+	–	+	+	+
Pallis'84[33]	0%	–	–	+	+	+	+	+

Randall'19[27]	0%	+	+	+	+	+	+	+
Samuelsson'06[61]	0%	+	+	+	+	+	+	+
Simpson'21[39]	0% [b]	+	−	+	+	+	+	+
Steeg'12[28]	0%	+	−	+	+	+	+	+
Steeg'18[62]	<1%	+	−	+	+	+	+	+
Stefansson'12[63]	0%	+	−	+	+	+	+	+
Wang'85[42]	0%	−	−	−	+	+	+	+
Case-Control Studies								
Braucht'70[40]	N/A	−	−	−	+	+	+	+
Dahle'05[66]	N/A	+	−	−	+	+	+	+
Dean'67[41]	N/A	−	−	−	+	+	+	+
Farberow'74[67]	N/A	+	+	−	+	+	+	+
Frottier'08[30]	N/A	−	−	−	+	+	+	+
Frottier'09[31]	N/A	+	−	−	+	+	+	+
Van de Loo'70[19]	N/A	−	−	−	+	+	+	+

N/A= Not applicable because case-control design; Std = Standard

Judgement: X High − Some concerns + Low

[a] Quality assessment is based on the QUADAS-2

[b] Authors highlighted a theoretical risk that people could have died outside of the catchment area and these deaths would not have been identified by the study.

[c] The study was designed as prospective analysis, although the control population was selected using case-control methodology.

[d] The authors report that data on suicide outcomes and test results were only available for a subset of the original population.