Suicide Risk Screening Tools for Pediatric Patients: A Systematic Review of Test Accuracy

Nathan J. Lowry, BA, Pauline Goger, PhD, Maria Hands Ruz, BA, Fangfei Ye, BS, Christine B. Cha, PhD

abstract

CONTEXT: Health care settings have increasingly adopted universal suicide risk screening tools into nonpsychiatric pediatric care; however, a systematic review examining the accuracy of these tools does not yet exist.

OBJECTIVE: Identify and review research on the test accuracy of suicide risk screening tools for pediatric patients in nonpsychiatric medical settings.

DATA SOURCES: PubMed and PsycINFO were searched to identify peer-reviewed articles published before March 23, 2023.

STUDY SELECTION: Articles that quantified the accuracy of a suicide risk screening tool (eg, sensitivity, specificity) in a nonpsychiatric medical setting (eg, primary care, specialty care, inpatient or surgical units, or the emergency department) were included.

DATA EXTRACTION: A total of 13 studies were included in this review. Screening tool psychometric properties and study risk of bias were evaluated.

RESULTS: Sensitivity among individual studies ranged from 50% to 100%, and specificity ranged from 58.8% to 96%. Methodological quality was relatively varied, and applicability concerns were low. When stratifying results by screening tool, the Ask Suicide-Screening Questions and Computerized Adaptive Screen for Suicidal Youth had the most robust evidence base.

LIMITATIONS: Because of considerable study heterogeneity, a meta-analytic approach was deemed inappropriate. This prevented us from statistically testing for differences between identified screening tools.

CONCLUSIONS: The Ask Suicide-Screening Questions and Computerized Adaptive Screen for Suicidal Youth exhibit satisfactory test accuracy and appear promising for integration into clinical practice. Although initial findings are promising, additional research targeted at examining the accuracy of screening tools among diverse populations is needed to ensure the equity of screening efforts.

Department of Counseling and Clinical Psychology, Teachers College, Columbia University, New York, New York

Mr Lowry conceptualized and designed the study and interpreted data; Dr Goger interpreted data; Dr Cha conceptualized and designed the study, interpreted data, and provided supervision; and all authors drafted the initial manuscript, critically reviewed and revised the manuscript for important intellectual content, approved the final manuscript as submitted, and agree to be accountable for all aspects of the work.

DOI: https://doi.org/10.1542/peds.2023-064172

Accepted for publication Nov 28, 2023

Address correspondence to Nathan J. Lowry, BA, Department of Counseling and Clinical Psychology, Teachers College, Columbia University, 525 W 120th St, New York, NY 10027. E-mail: njl2146@tc.columbia.edu

To cite: Lowry NJ, Goger P, Hands Ruz M, et al. Suicide Risk Screening Tools for Pediatric Patients: A Systematic Review of Test Accuracy. *Pediatrics*. 2024;153(3):e2023064172

Youth suicide is a prevalent and complex public health problem.[1] Moreover, it is resource- and time-intensive to effectively assess, predict, and treat. Universal suicide risk screening (ie, screening all pediatric patients for suicide risk regardless of presenting problem) is one way to feasibly identify and manage pediatric quickly "rule out" those who do not require further assessment, allowing medical settings to suicide risk.[2–6]

The practice of universal suicide risk screening can help guide medical providers on where to direct their clinical attention and resources. However, validated tools are needed to ensure that screening is both accurate and feasible. Emerging research suggests that single item screens are often inaccurate, resulting in both under- and overdetection of suicidality.[7–10] Missing youth at risk or overburdening limited mental health resources can directly result in negative clinical outcomes. Thus, it is important that tools are not just evidence-based, but also validated among the populations that are being screened.

Previous reviews[11,12] have identified evidence-based screening tools for pediatric patients, however, a formal synthesis of research examining the accuracy (ie, the ability of a test to correctly identify the presence or absence of suicidality) of these tools does not exist. Moreover, it is unclear how these tools perform in nonpsychiatric medical settings, where suicidality is not as prevalent or may not be the primary concern. Recently, medical organizations such as the American Academy of Pediatrics have recommended suicide risk screening for all pediatric patients ages 12 and older.[13] Thus, as nonpsychiatric health care settings increasingly adopt universal suicide risk screening tools into practice, it is important to evaluate their accuracy. To address this gap, this systematic review aims to (1) identify research examining the accuracy of suicide risk screening tools for pediatric patients in nonpsychiatric medical settings, (2) summarize psychometric properties of suicide risk screening tools, and (3) identify areas for future research.

METHODS

Search Strategy

This review follows Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA)[14] reporting guidelines and was preregistered with the International Prospective Register of Systematic Reviews (CRD42023406150).

Medline/PubMed and PsycINFO/EBSCO were searched to identify peer-reviewed articles published before March 23, 2023. The following search terms were used: (((“Adolescent”[Mesh]) OR “Child”[Mesh] OR “Youth” OR “Pediatric”) AND Suicid* AND (Screen* OR “Risk Assessment”[Mesh])). Two authors (M.H.-R. and F.Y.) independently screened abstracts for eligibility. Disagreements were resolved through the consensus of a third blind reviewer (N.L.). This process was repeated when screening full-text articles for eligibility. Screening was completed using the online platform, Covidence.

Inclusion Criteria

Eligible empirical articles met the following inclusion criteria: (1) written in English; (2) published in a peer-reviewed journal; (3) sample is youth aged 25 years old or younger; (4) study was conducted in a nonpsychiatric medical setting (eg, primary care, specialty care, inpatient or surgical units, or the emergency department [ED]); and (5) outcomes quantified the accuracy of a suicide risk screening tool (eg, sensitivity, specificity, positive predictive value [PPV], negative predictive value [NPV], area under the curve [AUC]). Articles were not excluded based on year published.

Data Extraction and Quality Assessment

Two authors (M.H.-R. and F.Y.) completed data extraction; data from all included articles was double coded to ensure accuracy. The following information was extracted from each article: (1) sample characteristics; (2) study setting; (3) screening tool characteristics; and (4) screening tool psychometric properties. If studies stratified results by nonpsychiatric and psychiatric medical settings, sample characteristics and results were extracted only from the nonpsychiatric population.

The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2)[15] tool was used to evaluate study risk of bias. QUADAS-2 signaling questions were adapted to evaluate potential bias in participant selection, index test, reference standard, and study flow and timing. We classified studies that lost at least 20% of participants to follow up as having “high” risk of bias.[16] Additionally, applicability of patient selection, index test, and reference standard was evaluated. Using the signaling questions, each study received a rating of low, high, or unclear for risk of bias and applicability. One author (P.G.) evaluated all included articles using the predetermined QUADAS-2 rating criteria. Articles that required additional consensus were evaluated by 2 authors (N.L. and C.C.) using the same QUADAS-2 criteria.

Data Synthesis

Extracted information was summarized using a narrative synthesis approach, with a focus on summary measures that quantify the accuracy of a suicide risk screening tools (eg, sensitivity, specificity, PPV, NPV, AUC).

RESULTS

Search Results

A total of 3560 articles were identified by our search. The oldest article identified by our search was published

in 1970. Given our broad search criteria, a total of 3520 were deemed ineligible at the title and abstract screening phase. A total of 40 full text articles were screened; 13 studies[17–29] met eligibility criteria and were included in this systematic review. Figure 1 provides a PRISMA diagram for the study selection process. Inter-rater agreement was moderate for the abstract screening phase ($\kappa = 0.42$) and almost perfect for the full-text screening phase ($\kappa = 0.82$).[30]

Test Accuracy and Sample Characteristics Across Studies

Sensitivity among individual studies ranged from 50% to 100% and specificity ranged from 58.8% to 96% (Table 1). For studies that reported AUC, values ranged from 0.754 to 0.956.

Across the 13 included studies, participant ages ranged from 8 to 22 years old (Table 2). Study samples were primarily female, white, and non-Hispanic. Sample sizes also varied considerably, ranging from 100 to 13 420.

Ask Suicide-Screening Questions

Eight studies identified by this review examined the accuracy of the Ask Suicide-Screening Questions (ASQ).[17–19,21,23–25,29] The ASQ is a 4-item self-report screening tool developed to identify clinically significant suicidal ideation or past behavior. The ASQ was first developed and validated among a sample of youth aged 10 to 21 presenting to the ED: when compared with the reference standard Suicidal Ideation Questionnaire (SIQ/SIQ-JR),[31,32] the ASQ had a sensitivity of 96.9% and specificity of 87.6%.[23] Subsequent research has examined the ASQ's psychometric properties in pediatric medical and surgical inpatient units, outpatient specialty care, and outpatient primary care. Using the SIQ/SIQ-JR as the reference standard, the ASQ demonstrated a sensitivity of 96.7% and specificity of 91.1% among medical and surgical inpatients aged 10 to 21.[24] In outpatient specialty care, the sensitivity was 100% and specificity 91.2% and in outpatient primary care, 100% and 87.9%.[17]

Additional research has evaluated the ASQ's predictive validity. One retrospective cohort study[19] reviewed medical records to determine the association between the ASQ and a suicide-related ED visit or death 3 months after screening. In this study, the ASQ yielded a sensitivity of 60% and specificity of 92.7%.[19] A separate retrospective cohort study[21] examined medical records of pediatric patients from the ED to determine how accurate the ASQ was in identifying individuals with a suicide-related

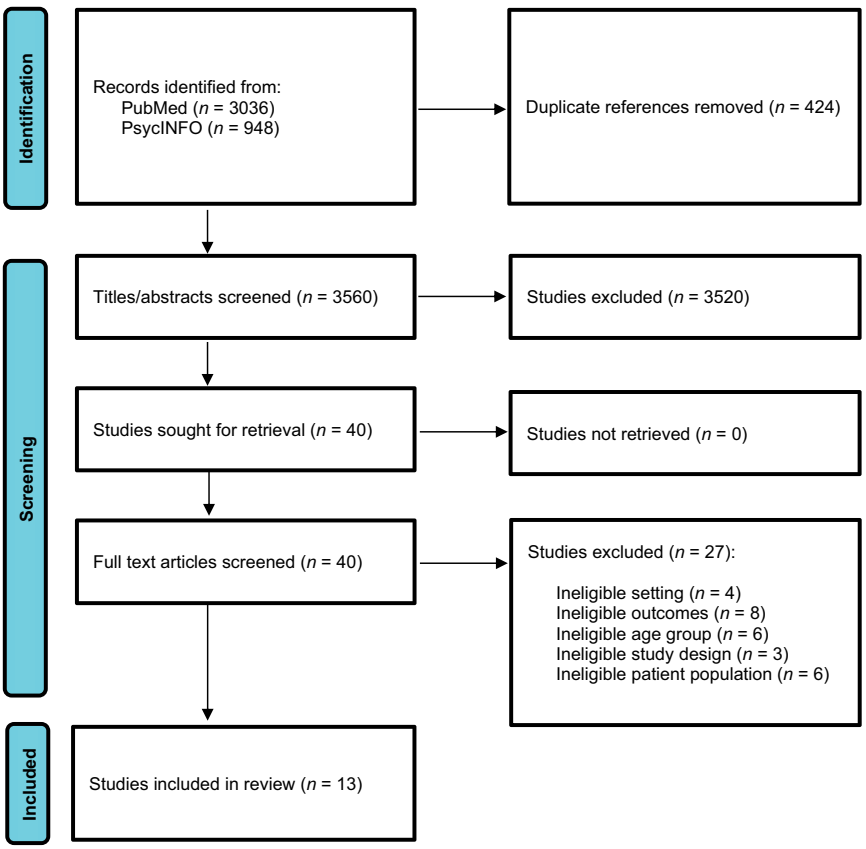

FIGURE 1
PRISMA flow diagram.

TABLE 1 Study Characteristics and Diagnostic Test Accuracy

Study	Screening Tool	Study Design	Setting	Reference Standard	Psychometric Properties				
					Sensitivity (95% CI)	Specificity (95% CI)	NPV (95% CI)	PPV (95% CI)	AUC
Aguinaldo et al (2021)[17]	ASQ	Cross sectional	Outpatient specialty care	SIQ/SIQ-JR	100 (80.5–100)	91.2 (87.5–94.1)	100 (98.7–100)	37.8 (23.8–53.5)	0.956
	ASQ	Cross sectional	Outpatient primary care	SIQ/SIQ-JR	100 (59.0–100)	87.9 (82.0–92.3)	100 (97.7–100)	25.0 (10.7–44.9)	0.939
Brent et al (2023)[18]	ASQ	Prospective	ER	SA at 3-mo follow up	95.1 (91.8–98.4)	58.8 (56.9–60.7)	99.5 (99.1–99.8)	12.7 (10.9–14.6)	0.769
	CASSY	Prospective	ER	SA at 3-mo follow up	94.5 (91.0–98.0)	64.3 (62.5–66.2)	99.5 (99.1–99.8)	14.4 (12.3–16.5)	0.867
DeVylder et al (2019)[19]	ASQ	Retrospective	ER	SA at 3-mo follow up	60.0 (44.2–74.3)	92.7 (92.1–93.2)	99.8 (99.7–99.9)	3.9 (3.1–5.0)	[a]
Diamond et al (2010)[20]	BHS	Cross sectional	Outpatient primary care	SSI	83.0 (71.0–90.0)	87.0 (83.0–91.0)	[a]	[a]	[a]
Haroz et al (2021)[21]	ASQ	Retrospective	ER	SA at 3-mo follow up	66.7	84.2	[a]	4.2	0.754
Hopper et al (2012)[22]	RSQ	Cross sectional	ER	SIQ/SIQ-JR	50.0 (1.0–99.0)	79.0 (69.0–86.0)	99.0 (93.0–100)	5.0 (0.0–23.0)	[a]
Horowitz et al (2012)[23]	ASQ	Cross sectional	ER	SIQ/SIQ-JR	96.9 (91.3–	87.6 (84.0–90.5)	[a]	[a]	0.920
Horowitz et al (2020)[24]	ASQ	Cross sectional	Medical or surgical inpatient unit	SIQ/SIQ-JR	96.7 (82.8–99.2)	91.1 (88.4–93.3)	99.8 (98.9–99.9)	36.3 (25.8–47.8)	0.940
Horowitz et al (2021)[26]	PHQ-A Item 9	Cross sectional	Medical or surgical inpatient unit	SIQ/SIQ-JR	70.0 (51.0–85.0)	96.0 (94.0–98.0)	[a]	[a]	[a]
Horowitz et al (2022)[6]	ASQ	Cross sectional	Med, surg, ER, outpatient	SIQ/SIQ-JR	92.0 (84.5–99.5)	91.7 (90.0–93.4)	[a]	[a]	0.918
King et al (2021)[27]	CASSY	Prospective	ER	SA at 3-mo follow up	82.4	72.5	[a]	[a]	0.870
Moss et al (2022)[28]	PHQ-9 Item 9	Cross sectional	Outpatient specialty care	C-SSRS	53.3 (27.4–77.7)	95.7 (89.9–98.4)	94.2 (87.9–97.4)	61.5 (32.3–84.9)	[a]
Uygun et al (2022)[29]	ASQ	Cross sectional	ER	SPS	100 (100–100)	75.3 (66.3–84.2)	100 (100–100)	4.05 (2.82–5.81)	0.876

C-SSRS, Columbia Suicide Severity Rating Scale; RSQ, Risk of Suicide Questionnaire; SPS, Suicide Probability Scale; SA, Suicide Attempt; SSI, Scale for Suicidal Ideation.
[a] Statistic not reported.

ED visit 3 months after screening. This study found the ASQ to have a sensitivity of 66.7% and specificity of 84.2%.[21] Lastly, in 1 prospective study of pediatric ED patients,[18] the ASQ had a sensitivity of 95.1% and specificity of 58.8% when using a suicide attempt 3 months post screening as the reference standard.

Two studies[25,29] examined the cross cultural validity of the ASQ. One examined the validity of a Turkish language version of the ASQ for screening in a Turkish ED. Using the Suicide Probability Scale[33] as the reference standard, the Turkish ASQ had a sensitivity of 100% and specificity of 75.3%. The second study examined the accuracy of the ASQ among Black youth.[25] This was a secondary analysis using data from 3 previous ASQ studies[17,23,24] and found no significant differences in ASQ psychometric properties between Black youth (sensitivity: 94%, specificity: 91.4%) and white youth (sensitivity: 90.9%, specificity: 91.8%).[25]

Behavioral Health Screen

One study[20] examined the accuracy of the Behavioral Health Screen (BHS), a self-report screening tool designed for use in outpatient primary care settings. This internet-based tool encompasses 13 domains ranging from nutrition to anxiety. One module within the BHS is the "suicide and self-harm" subscale, which is comprised of 5 core items and 5 follow-up items. This module identifies past week and lifetime suicide risk among adolescents and young adults. In its original validation study, the BHS "suicide and self-harm" subscale demonstrated a sensitivity of 83% and specificity of 87%[20] when compared with the Scale for Suicidal Ideation.[34]

Computerized Adaptive Screen for Suicidal Youth

Two studies[18,27] examined the accuracy of the Computerized Adaptive Screen for Suicidal Youth (CASSY), an electronic self-report screening tool designed to identify youth at risk for future suicidal behavior. The CASSY is a computerized

TABLE 2 Study Sample Characteristics

Study	*N*	Age Range	Age Mean (SD)	Sex (% Female)	Race	Ethnicity (% Hispanic)
Aguinaldo et al (2021)[17]	515	10–21	15.3 (2.7)	52.8	White: 50.1%; Black: 18.4%; Hispanic: 20.4%; Asian: 7%; unknown: 4.1%	20.4
Brent et al (2023)[18]	2740	12–17	15.0 (1.7)	62.2	White: 59.1%; Black or African American: 17.1%; unknown or unavailable: 11.9%; multiracial: 5.9%; American Indian or Alaska Native: 3.8%; Asian, Native Hawaiian or other Pacific Islander: 2.3%	24.7
DeVylder et al (2019)[19]	10 337	8–18	14.7 (3.2)	52.1	Black: 67.4%; white: 21.8%; Asian: 1.4%; Latino: 6.1%; other: 3.3%	6.14
Diamond et al (2010)[20]	415	12–21	15.8 (2.2)	66.5	Black: 77.5%; white: 10.7%; Hispanic: 9.7%; other: 2.1%	9.7
Haroz et al (2021)[21]	13 420	8–18	14.3	52.5	Black: 63%; white: 25%; Hispanic: 25%; multiracial: 5.2%	6.8
Hopper et al (2012)[22,a]	100	13–18	14.5 (1.1)	40	[b]	[b]
Horowitz et al (2012)[23]	524	10–21	15.2 (2.6)	56.9	White 50.4%; Black 29.6%; Hispanic/Latino 9.0%; Asian 2.3%; other or unknown 8.8%	9.0
Horowitz et al (2020)[24]	600	10–21	15.4 (2.8)	59.17	White 55.17%; African American 23.33%; Hispanic/Latino 9.83%; multiple races 7.5%; Asian 1.66%; other 1.83%; unknown 0.66%	9.83
Horowitz et al (2021)[26]	600	10–21	15.4 (2.8)	59.17	White 55.17%; African American 23.33%; Hispanic/Latino 9.83%; multiple races 7.5%; Asian 1.66%; other 1.83%; unknown 0.66%	9.83
Horowitz et al (2022)[6,a]	1083	10–21	[b]	56.2	Non-Hispanic white 51.8% Non-Hispanic Black 22.7%	0
King et al (2021)[27]	2754	12–17	15.1 (1.61)	63	White: 54%; Black or African American: 23%; unknown or unavailable: 13%; multiracial: 6%; Asian, Native Hawaiian or other Pacific Islander: 2%; American Indian or Alaska Native: 1%	23
Moss et al (2022)[28]	133	16–22	19.6 (1.1)	51.1	White: 73.3%; Black: 13.3%; Asian: 6.7%; other: 6.7%	9.0
Uygun et al (2022)[29,a]	100	10–18	14.7 (1.9)	52.4	[b]	[b]

[a] Hopper et al (2012) enrolled a sample of Australian youth. Horowitz et al (2022) is a secondary analysis that examined non-Hispanic Black and non-Hispanic white youth. Uygun et al (2022) enrolled a sample of Turkish youth.
[b] Statistic not reported.

adaptive test, such that an algorithm determines what items are administered based on how the respondent answers. These questions are drawn from a pool of 72 items that cover various domains, including post traumatic stress disorder, social adjustment, sleep, substance use, anger, and aggression. On average, the CASSY administers 11 items (range 5 to 21). Of note, the CASSY includes 3 of the 4 ASQ items in every administration.

The CASSY was originally developed utilizing prospective data from pediatric EDs and then independently validated with a sample pediatric ED patients 12 to 17-year-olds.[27] In this study, the CASSY had a sensitivity of 82.4% and specificity of 72.5% when compared with the reference standard (a self-reported suicide attempt 3 months after screening).[27] Another study[18] utilizing a sample of pediatric ED patients extended this research and found the CASSY to have a sensitivity of 94.5% and specificity of 64.3%.

Patient Health Questionnaire-9, Item 9

Two studies[26,28] examined the accuracy of the Patient Health Questionnaire-9 (PHQ-9). The PHQ-9 is a self-report depression screening tool that evaluates depressive symptoms over the past 2 weeks and includes 1 item (Item 9) that screens for thoughts of suicide and self-harm.[35] One study[28] examined the psychometric properties of the PHQ-9 Item 9 in an outpatient specialty care unit for 12 to 22-year-olds by using the Columbia-Suicide Severity Rating Scale[36] as the reference standard. In this cross-sectional study, PHQ-9 Item 9 had a sensitivity of 53.3% and specificity of 95.7%.[28]

Another study identified by this review assessed the validity of the Patient Health Questionnaire-Adolescent Version (PHQ-A) Item 9.[26] This study used a convenience sample of medical and surgical inpatients aged 10 to 21-year-old. Using the SIQ/SIQ-JR[31] as the reference

standard, PHQ-A Item 9 had a sensitivity of 70.0% and specificity of 96.0%.[26]

Risk of Suicide Questionnaire

One study[22] investigated the accuracy of the Risk of Suicide Questionnaire (RSQ), a 4-item verbally administered screening tool that identifies current suicidality. This study used a sample of pediatric patients ages 13 to 18 without mental health complaints or recent psychiatric history who presented to the ED.[22] Of note, none of the participants who screened positive on the RSQ also screened positive on the reference standard SIQ (cutoff score >41) or SIQ-JR (cutoff score >31). When using more liberal cutoff scores of 30 on the SIQ and 23 on the SIQ-JR, the RSQ

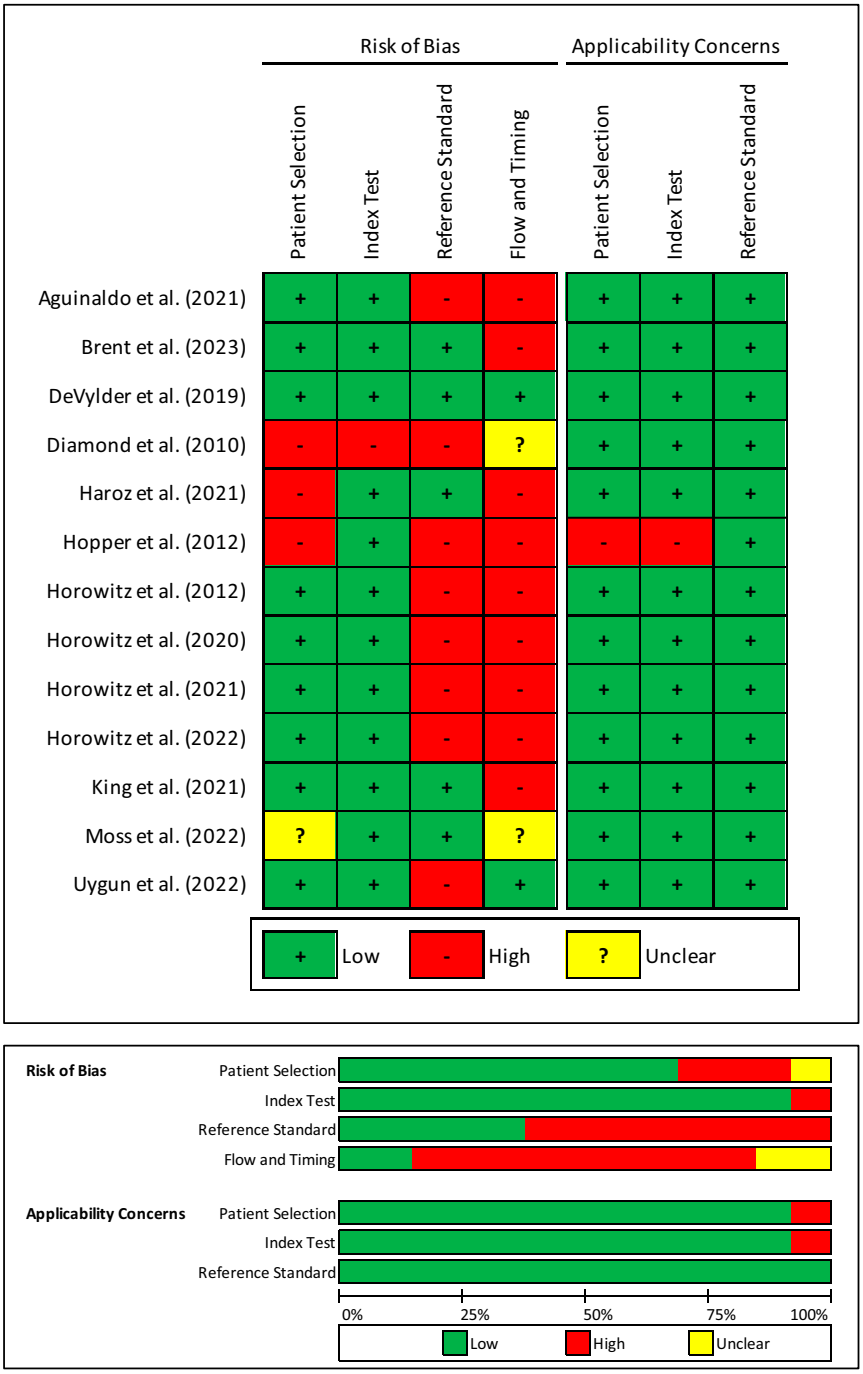

FIGURE 2
QUADAS-2 ratings of included studies.

demonstrated a sensitivity of 50% and specificity of 79%.[22]

Methodological Quality of Included Studies

Methodological quality was relatively varied across domains (Fig 2). The index test and patient selection domains exhibited a low risk of bias, apart from 1 and 3 studies rated as having high bias in these categories, respectively. In contrast, at least 50% of studies were classified as having high risk of bias for the reference standard and flow and timing domains, mostly because of using survey measures as reference standard and not administering the same version of the reference standard to all participants. Applicability concerns were very low, with hardly any studies showing bias across these domains.

DISCUSSION

The findings of this review demonstrate that certain screening tools for suicide risk exhibit satisfactory accuracy and hold promise for integration into clinical practice. We provide commentary on the accuracy of identified tools, where study methodology can be improved, considerations for the cross-cultural validity of screening tools, the role of screening within clinical pathways, and optimal clinical targets for youth risk screening.

Test Accuracy of Identified Screening Tools

The ASQ was the most widely studied tool in this review and demonstrated strong psychometric properties. Additionally, it was the only tool to be evaluated and validated for use among pediatric patients in the ED, inpatient and medical surgical units, and outpatient primary and specialty care. The newly developed CASSY also shows promise as a screening tool, demonstrating strong sensitivity and specificity for use in the ED while leveraging novel adaptive testing methods. However, the accuracy of the CASSY in other medical settings (eg, outpatient primary care) has yet to be examined and is an area for future research to address.

The BHS displayed appropriate accuracy in outpatient primary care. However, it should be noted that the BHS is a broadband measure that screens for other psychopathology and risky behavior, and it is unclear how its suicide risk subscale would perform in isolation. Moreover, further research examining the accuracy of the BHS in other medical settings is needed to determine its utility in other medical settings.

Our findings suggest that caution should be exercised when using certain screening tools. Both the PHQ-9/PHQ-A, Item 9, and the RSQ demonstrated poor sensitivity in their respective studies. This may in part result from the fact that neither tool was specifically developed for universal suicide risk screening. The PHQ has been validated for use as a depression screening tool,[37] however, its reliance on a single item to screen for suicide risk may make it susceptible to missing patients who warrant further examination. Similarly, the RSQ was originally developed to identify suicide risk among psychiatric inpatients, which may contribute to its psychometric properties not generalizing to a nonpsychiatric medical population. Additionally, in its original validation study, the RSQ demonstrated a sensitivity of 98% and specificity of 37%,[38] suggesting that it may also be prone to generating a false positives.

Areas for Methodological Improvement

Using the QUADAS-2,[15] this review identified areas where the methodological quality of suicide screening research can be strengthened to avoid risk of bias, specifically in the domains of reference standard and flow and timing. Many studies[17,20,22–26,29] opted to use a self-report survey measure as the reference standard (eg, the SIQ or Scale for Suicidal Ideation), which inherently introduced the possibility of measurement error and risk of bias. This may have resulted in the over- or underestimation of screening tool accuracy. Future research should aspire to use clinician assessment or behavioral outcomes (eg, suicide attempts) as the "gold standard" when validating tools.

Flow and timing of administration represented a significant area in need of improvement across studies. Several studies[17,22–26] administered a different reference standard to participants based on age (eg, the SIQ and SIQ-JR). Administering either the SIQ or SIQ-JR was necessary when choosing the SIQ because these tools are validated for different age groups,[31,32] however, this is not considered best practice when determining the diagnostic test accuracy of a tool[15] and other options (eg, clinician assessment) were available and would have addressed this issue. Other studies suffered from participant attrition either because of losing more than 20% of participants to follow-up[18,27] or by retroactively selecting only complete screening cases from medical records.[19,21] Participant retention is often a challenge in longitudinal studies, though the potential for misclassification resulting in the overestimation of sensitivity and specificity should not be overlooked. Lastly, 1 study[21] reported that the index test was not administered with fidelity.

Accuracy of Tools among Diverse Populations

There is a clear need for additional research to examine the accuracy of suicide risk screening tools among diverse youth populations. This is a particularly relevant area of study as cultural differences can influence how an individual perceives and discusses suicidality.[39] Overall, samples across studies were primarily female, white, and non-Hispanic (Table 2). Moreover, only 1 study directly examined the validity of screening among youth of color (ie, Black youth),[25] raising concerns about the generalizability of findings.

Translations of screening tools are also needed to account for the diversity of the youth that present during screening. Moreover, it is crucial to validate these translations to ensure that tools have relevant vernacular and retain the intended meaning of questions. One identified study validated a Turkish language version of the ASQ.[29] Other research that did not meet inclusion criteria has examined the validity of screening tool translations[40,41] and current research is underway to validate translations of screening tools.[42] Nevertheless, further research in this area is needed to ensure the equity of screening.

Clinical Pathways and the Role of Screening

Screening tool accuracy should also be evaluated in the context of larger clinical pathways. The American Academy of Pediatrics (AAP) provides guidance on youth suicide prevention strategies in clinical settings.[43] Referencing clinical pathways developed by an American Academy of Child and Adolescent Psychiatry workgroup,[44] AAP emphasizes screening as the first step in a 3-tiered pathway. Within these pathways, screening tool sensitivity is valued over specificity, which may result in false positives. To address the issue, patients who screen positive on the initial suicide risk screen receive a brief suicide safety assessment, which acts as a triage step to confirm the presence of suicidality. Using information collected during the brief suicide safety assessment, a provider can then determine appropriate safety precautions and next steps for care (eg, a full mental health evaluation, outpatient referral, or no action). Thus, lower screening tool specificity values may be tolerable. However, there is currently no standard of care for managing suicide risk in the medical setting, and the pathways highlighted by AAP represent one possible approach. Future research aimed at studying these pathways would help clarify the role of screening and the necessary accuracy to optimize the feasibility of risk identification.

Identifying an Optimal Reference Standard

One important discrepancy between studies identified by this review was the reference standard used for validation. Although all the tools identified by this review screen for "suicide risk," some studies used a reference standard of current suicidal ideation, (ie, the SIQ) whereas others used future suicidal behavior (ie, suicide attempts). This difference is representative of current debate on the optimal target of suicide risk screening.[45–47] Notably, the use of different reference standards may have important implications for the accuracy of suicide risk screening tools and the clinical information these tools provide to clinicians.

Suicide attempts are an important outcome to prevent, however, future suicidal behavior is notoriously difficult to predict,[48] which may result in poor suicide risk screening tool accuracy. Instead, a de-emphasis on future behavior and focus on detecting clinically significant suicidal ideation may improve the accuracy of screening tools. Moreover, although suicidal ideation is a poor predictor of suicide mortality,[49,50] identifying youth with suicidal thoughts has important clinical utility, regardless of whether they go on to die by suicide. Compared with adolescents without suicidal ideation, adolescents with suicidal ideation are significantly more likely to develop psychopathology and have poor overall functioning by age 30.[51] Thus, suicidal ideation appears to be an important clinical indicator of distress that is relevant to the detection of youth suicidality and overall well-being. However, this is an emerging area of research, and future studies should investigate how different reference standards impact suicide risk screening tool accuracy and clinical outcomes.

Limitations

The findings of this review should be interpreted with the following limitations. Foremost, despite a comprehensive search, it is possible that articles relevant to this review were missed. Additionally, because of considerable study heterogeneity, a meta-analytic approach was deemed inappropriate. This prevented us from statistically testing for differences between identified screening tools.

CONCLUSIONS

The findings of this review demonstrate that screening tools for suicide risk exhibit satisfactory test accuracy and hold promise for integration into clinical practice. However, practicing caution is necessary when choosing screening tools, as some commonly recommended tools lack sufficient research to support their validity for pediatric patients and require further evaluation. Additionally, the low number of studies identified by this review suggests that this is a novel field of research that requires further study. Future research targeted at studying the accuracy of screening tools among diverse populations is particularly needed to ensure the equity of universal suicide risk screening efforts.

ABBREVIATIONS

ASQ: Ask Suicide-Screening Questions
BHS: Behavioral Health Screen
CASSY: Computerized Adaptive Screen for Suicidal Youth
PHQ-9: Patient Health Questionnaire
PHQ-A: Patient Health Questionnaire, Adolescent Version
RSQ: Risk of Suicide Questionnaire
SIQ: Suicidal Ideation Questionnaire
SIQ-JR: Suicidal Ideation Questionnaire, Junior Version

FUNDING: No external funding.

CONFLICT OF INTEREST DISCLOSURES: The authors have indicated they have no conflicts of interest to report.

REFERENCES

1. Centers for Disease Control and Prevention. Web-based injury statistics query and reporting system (WISQARS). Available at: https://www.cdc.gov/injury/wisqars. Accessed August 31, 2022
2. Hughes JL, Horowitz LM, Ackerman JP, Adrian MC, Campo JV, Bridge JA. Suicide in young people: screening, risk assessment, and intervention. *BMJ*. 2023:381
3. Roaten K, Horowitz LM, Bridge JA, et al. Universal pediatric suicide risk screening in a health care system: 90,000 patient encounters. *J Acad Consult Liaison Psychiatry*. 2021;62(4):421–429
4. Sullivant SA, Brookstein D, Camerer M, et al. Implementing universal suicide risk screening in a pediatric hospital. *Jt Comm J Qual Patient Saf*. 2021;47(8):496–502
5. Lois BH, Urban TH, Wong C, et al. Integrating suicide risk screening into pediatric ambulatory subspecialty care. *Pediatr Qual Saf*. 2020;5(3):e310
6. Horowitz LM, Bridge JA, Tipton MV, et al. Implementing suicide risk screening in a pediatric primary care setting: from research to practice. *Acad Pediatr*. 2022;22(2):217–226
7. Ammerman BA, Burke TA, Jacobucci R, McClure K. How we ask matters: The impact of question wording in single-item measurement of suicidal thoughts and behaviors. *Prev Med*. 2021;152(Pt 1):106472
8. Gratch I, Tezanos KM, Fernades SN, Bell KA, Pollak OH, Cha CB. Single- vs. multi-item assessment of suicidal ideation among adolescents. *R I Med J*. 2022;105(4):16–21
9. Hom MA, Joiner TE, Bernert RA. Limitations of a single-item assessment of suicide attempt history: implications for standardized suicide risk assessment. *Psychol Assess*. 2016;28(8):1026–1030
10. Millner AJ, Lee MD, Nock MK. Single-item measurement of suicidal behaviors: validity and consequences of misclassification. *PLoS One*. 2015;10(10):e0141606
11. Thom R, Hogan C, Hazen E. Suicide risk screening in the hospital setting: a review of brief validated tools. *Psychosomatics*. 2020;61(1):1–7
12. Scudder A, Rosin R, Baltich Nelson B, Boudreaux ED, Larkin C. Suicide screening tools for pediatric emergency department patients: a systematic review. *Front Psychiatry*. 2022;13:916731
13. Health Resources and Services Administration. Updates to the Bright Futures periodicity schedule. Available at: https://www.federalregister.gov/documents/2022/01/12/2022-00461/updates-to-the-bright-futures-periodicity-schedule. Accessed December 11, 2022
14. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71
15. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529–536
16. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet*. 2002;359(9308):781–785
17. Aguinaldo LD, Sullivant S, Lanzillo EC, et al. Validation of the ask suicide-screening questions (ASQ) with youth in outpatient specialty and primary care clinics. *Gen Hosp Psychiatry*. 2021;68:52–58
18. Brent DA, Horowitz LM, Grupp-Phelan J, et al; Pediatric Emergency Care Applied Research Network (PECARN). Prediction of suicide attempts and suicide-related events among adolescents seen in emergency departments. *JAMA Netw Open*. 2023;6(2):e2255986
19. DeVylder JE, Ryan TC, Cwik M, et al. Assessment of selective and universal screening for suicide risk in a pediatric emergency department. *JAMA Netw Open*. 2019;2(10):e1914070
20. Diamond G, Levy S, Bevans KB, et al. Development, validation, and utility of internet-based, behavioral health screen for adolescents. *Pediatrics*. 2010;126(1):e163–e170
21. Haroz EE, Kitchen C, Nestadt PS, Wilcox HC, DeVylder JE, Kharrazi H. 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*. 2021;51(6):1189–1202
22. Hopper SM, Woo JW, Sharwood LN, Babl FE, Long EJ. Prevalence of suicidality in asymptomatic adolescents in the paediatric emergency department and utility of a screening tool. *Emerg Med Australas*. 2012;24(5):540–546
23. Horowitz LM, Bridge JA, Teach SJ, et al. Ask suicide-screening questions (ASQ): a brief instrument for the pediatric emergency department. *Arch Pediatr Adolesc Med*. 2012;166(12):1170–1176
24. Horowitz LM, Wharff EA, Mournet AM, et al. Validation and feasibility of the ASQ among pediatric medical and surgical inpatients. *Hosp Pediatr*. 2020;10(9):750–757
25. Horowitz LM, Mournet AM, Sheftall A, et al. Assessing the validity of the ask suicide-screening questions in black youth. *J Acad Consult-Liaison Psychiatry*. 2023;64(4):332–335
26. Horowitz LM, Mournet AM, Lanzillo E, et al. Screening pediatric medical patients for suicide risk: is depression screening enough?. *J Adolesc Health*. 2021;68(6):1183–1188
27. King CA, Brent D, Grupp-Phelan J, et al; Pediatric Emergency Care Applied Research Network. Prospective development and validation of the computerized adaptive screen for suicidal youth. *JAMA Psychiatry*. 2021;78(5):540–549

28. Moss AC, Roberts AJ, Yi-Frazier JP, et al. Identifying suicide risk in adolescents and young adults with type 1 diabetes: are depression screeners sufficient? *Diabetes Care*. 2022;45(5):1288–1291

29. Uygun SD, Çıkılı Uytun M, Konşuk Ünlü H, et al. Turkish adaptation study of the ask suicide-screening questions in adolescents. *Noro Psikiyatr Ars*. 2022;60(2):158–168

30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174

31. Reynolds WM, Mazza JJ. Assessment of suicidal ideation in inner-city children and young adolescents: reliability and validity of the Suicidal Ideation Questionnaire-JR. *School Psych Rev*. 1999;28:17–30

32. Reynolds WM. *SIQ, Suicidal Ideation Questionnaire*. Psychological Assessment Resources; 1988

33. Atli Z, Eskin M, Dereboy C. The validity and the reliability of suicide probability scale (SPS) in clinical sample. Turk Psikiyatri Derg. 2009;12:111–124

34. Beck AT, Brown GK, Steer RA. Psychometric characteristics of the Scale for Suicide Ideation with psychiatric outpatients. *Behav Res Ther*. 1997;35(11):1039–1046

35. Johnson JG, Harris ES, Spitzer RL, Williams JBW. The patient health questionnaire for adolescents: validation of an instrument for the assessment of mental disorders among adolescent primary care patients. *J Adolesc Health*. 2002;30(3):196–204

36. Posner K, Brown GK, Stanley B, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *Am J Psychiatry*. 2011;168(12):1266–1277

37. Costantini L, Pasquarella C, Odone A, et al. Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): a systematic review. *J Affect Disord*. 2021;279:473–483

38. Horowitz LM, Wang PS, Koocher GP, et al. Detecting suicide risk in a pediatric emergency department: development of a brief screening tool. *Pediatrics*. 2001;107(5):1133–1137

39. Molock SD, Boyd RC, Alvarez K, et al. Culturally responsive assessment of suicidal thoughts and behaviors in youth of color. *Am Psychol*. 2023;78(7):842–855

40. Robles-Garcia R, Páez Agraz F, Ascensio Guirado M, Mercado Salcedo E, Hernández Muñoz L. Evaluation of suicide risk in children: psychometric properties of the Spanish version of the Risk of Suicide Questionnaire (RSQ). *Actas Esp Psiquiatr*. 2005;33(5):292–297

41. Kim HJ, Kim S, Son Y, Youn I, Lee K. Reliability and validity of the Korean Version of the Ask Suicide-Screening Questions (ASQ). *J Korean Med Sci*. 2023;38(6):e41

42. Horowitz LM, Lowry NJ. *Adapting suicide risk screening for global pediatric medical settings: Do questions about suicide translate?* [Symposium]. Barcelona, Spain: IASR/AFSP International Summit on Suicide Research; 2023

43. American Academy of Pediatrics. Blueprint for youth suicide prevention. Available at: https://www.aap.org/suicideprevention. Accessed August 31, 2022

44. Brahmbhatt K, Kurtz BP, Afzal KI, et al; PaCC Workgroup. Suicide risk screening in pediatric hospitals: clinical pathways to address a global health crisis. *Psychosomatics*. 2019;60(1):1–9

45. Nestadt PS, Triplett P, Mojtabai R, Berman AL. Universal screening may not prevent suicide. *Gen Hosp Psychiatry*. 2020;63:14–15

46. Obegi JH. Is identifying suicide risk the point of universal screening? *Gen Hosp Psychiatry*. 2019;59:80

47. Simon GE, Richards JE, Whiteside U. Reframing the key questions regarding screening for suicide risk. *JAMA*. 2023;329(23):2026–2027

48. Nock MK, Millner AJ, Ross EL, et al. Prediction of suicide attempts using clinician assessment, patient self-report, and electronic health records. *JAMA Netw Open*. 2022;5(1):e2144373

49. Franklin JC, Ribeiro JD, Fox KR, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol Bull*. 2017;143(2):187–232

50. Ribeiro JD, Franklin JC, Fox KR, et al. Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies. *Psychol Med*. 2016;46(2):225–236

51. Reinherz HZ, Tanner JL, Berger SR, Beardslee WR, Fitzmaurice GM. Adolescent suicidal ideation as predictive of psychopathology, suicidal behavior, and compromised functioning at age 30. *Am J Psychiatry*. 2006;163(7):1226–1232