BMJ Open

Predicting future self-harm or suicide in adolescents: a systematic review of risk assessment scales/tools

Isobel Marion Harris,[ORCID] Sophie Beese, David Moore

Institute of Applied Health Research, University of Birmingham, Birmingham, UK

Correspondence to
Isobel Marion Harris;
i.m.harris@bham.ac.uk

ABSTRACT

Objective This systematic review aimed to evaluate the ability of risk tools to predict the future episodes of suicide/self-harm in adolescents.

Design Systematic review.

Data sources MEDLINE, EMBASE, CINAHL and PsycINFO were searched from inception to 3 March 2018.

Eligibility criteria for selecting studies Cohort studies, case–control studies and randomised controlled trials of adolescents aged 10–25 who had undergone risk assessment in a clinical setting following an episode of self-harm were included.

Data extraction and synthesis Two independent reviewers extracted data and assessed risk of bias. Data were grouped by tool and narrative synthesis undertaken, with studies appraised using a checklist combining the QUIPS (Quality In Prognosis Studies) and QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) tools.

Results Of the 17 137 articles initially identified, 11 studies evaluating 10 separate tools were included. The studies varied in setting, population and outcome measure. The majority of the studies were rated as having an unclear risk of bias, and meta-analysis was not possible due to high variability between studies.

The ability of the tools to correctly identify those adolescents going on to make a self-harm/suicide attempt ranged from 27% (95% CI 10.7% to 50.2%) to 95.8% (95% CI 78.9% to 99.9%). A variety of metrics were provided for 1–10 points increases in various tools, for example, odds and HRs.

Conclusions This systematic review is the first to explore the use of assessment tools in adolescents. The predictive ability of these tools varies greatly. No single tool is suitable for predicting a higher risk of suicide or self-harm in adolescent populations.

PROSPERO registration number CRD42017058686

INTRODUCTION

Self-harm and suicide are serious problems in children, adolescents and young people,[1–3] with the highest rates seen in those aged 16–24 years.[4,5] The term self-harm can be used to refer to acts of self-poisoning or self-injury (eg, cutting, scratching, breaking bones and burning) carried out intentionally regardless of motive or suicidal intent.[6] Repetition of self-harm is common, with 15%–25% of adolescents, treated at hospital for an

Strengths and limitations of this study

► This systematic review is the first to explore the use of tools to predict future self-harm/future suicide attempts in an adolescent population
► A checklist combining the QUIPS and QUADAS-2 tools was used to evaluate the quality of the included studies
► High levels of heterogeneity meant that meta-analysis was not possible
► Results of the study support other reviews in adult populations and highlights the need for further risk prediction work in this area

episode of self-harm returning for treatment within 12 months.[2,7]

There is a strong association between self-harm and risk of future suicide,[6,8] with approximately 50% of adolescents who die by suicide having previously self-harmed,[9] and self-harm increasing the risk of death by suicide approximately tenfold.[10] The 6-month period, following an episode of self-harm, has been identified as when this risk of death is highest.[11] Despite the association, there are important differences between acts of self-harm and suicide attempts. The motivations behind suicide attempts include wanting to end one's life, whereas motivations behind self-harm are more typically related to attenuating negative emotions and feelings.[12] There are also key differences with suicide attempts tending towards high lethality, infrequent occurrence and single method use only, whereas acts of self-harm contrast with single or multiple methods of low lethality and repetition.[13]

Predicting suicide and repetition of self-harm in adolescents is a challenge as it is usually a secretive and hidden behaviour, and there is no single risk factor for adolescent self-harm.[1,14–16] Additionally, in this population, suicide is a rare event with low prevalence (5.8 per 100 000 for the 15–19 years age group, and 9.5 per 100 000 for the 20–24

years age group in the UK in 2015[17] making risk prediction difficult.

The UK guidelines for the management of self-harm in those aged over 8 years recommend that all patients who present with self-harm should undergo comprehensive psychosocial assessment including assessing the risk of repetition of self-harm or suicide, as well as a full mental health and social needs assessment.[18–20] Risk scales/tools (from here on referred to a tools) tend to be a key part of this assessment, however, there is currently only a small amount of evidence available regarding their use and effectiveness, and no guidance as to which to use, or which is best for particular populations or settings.[18 21] The UK guidelines suggest that these risk tools should not be used alone to determine future risk or to make a decision on when to offer treatment.[18–20] There is a large variation in the type and format of risk tools being used across the UK.[22–24] A survey of 32 English hospitals found that over 20 different risk tools were in use, many of these were locally developed, highlighting a lack of consistency of practice.[21]

The content is variable from tool to tool with some only assessing a few parameters and others assessing a more extensive range. Adult tools may not be appropriate for use with adolescents due to either inappropriate questions being asked, or important areas specific to the age group not being assessed.

Previous systematic reviews of risk assessment tools used to predict future self-harm or suicide following self-harm in an adult population have been conducted,[3 25–27] and none were able to conclude that any one tool performed better than another or was more useful for predicting future self-harm or suicide attempts.

While adult populations have been considered, there is currently no systematic review examining risk tools in an adolescent population, a key age group affected by self-harm and suicide. A review is required specifically focusing on this age group to allow greater understanding of the use of risk assessments in this population.

The wide variety in the type of risk tools being used across hospitals and current lack of guidance surrounding their use in particular patient groups and settings warrants further investigation to improve and standardise patient care. This review contributes to the body of evidence for self-harm and suicide prediction and prevention to ensure that informed decisions can be made regarding predication of future risk and future research.

Objectives

This systematic review aimed to evaluate the ability of risk assessment tools to predict future episodes of self-harm or suicide in adolescents and young adults presenting to clinical services with an episode of self-harm or attempted suicide. To achieve this aim, the review asked the following question:

How accurately do risk assessment tools/scales predict which adolescent and young adults will go on to self-harm in the future, make a future suicide attempt or die by suicide?

METHODS
Search strategy

The following bibliographic databases were searched from inception to March 2018 with no language restrictions: MEDLINE, EMBASE, CINAHL, PsycINFO and Open Grey. Searches used index and free-text terms related to self-harm, suicide, adolescents and risk assessment. Terms were combined using the appropriate OR and AND operators (see MEDLINE search strategy in online supplementary appendix 1).

Selection/inclusion criteria
Study design

Prospective and retrospective cohort studies, case–control studies and randomised controlled trials testing assessment tools.

Patient group

Adolescents and young people aged 10–25 who have self-harmed or attempted suicide who have presented to clinical services or have been treated by a clinician following an episode of self-harm or attempted suicide. Studies also containing older populations were included if data for 10–25 years were presented separately. Studies on a wider age range were included if the majority of the sample (>50%) had self-harmed or attempted suicide.

WHO defines adolescence as being between 10 and 19 years[28], however, in the literature the upper limit for adolescents can range from 18 to 25 years.[1] In an attempt to include all relevant data in this review, reflecting this and that self-harm is rare before the age of 12, a broad range of 11–25 years will be included.[29]

Interventions

Any risk assessment.

Timing

No restriction was placed on length of follow-up so long as the outcomes occurred after the risk assessment was carried out.

Setting

Studies with the risk assessment carried out within a clinical setting or by a clinician were included. This could have been an inpatient or outpatient facility or as part of a home treatment programme.

Outcome

Self-harm and/or attempted suicide or completed suicide. It was acknowledged that these may be recorded individually or they may be grouped together (eg, repeat self-harm and attempted suicide may have been classed as the same event).

Study selection

Due to the large number of records, 10% of titles and abstracts were initially screened by two reviewers independently using predefined criteria based on the target

population and outcome to identify potentially relevant articles. After discussion and consideration of reasons for any disagreements, the remaining titles and abstracts were screened by one reviewer.

The full text of the potentially relevant articles was obtained and assessed against the full inclusion criteria. Endnote V.X7 software (Clarivate Analytics) was used to record study selection decisions, and reasons for exclusion were noted.

Data extraction strategy

Data extracted included: study characteristics (duration, start and end date, country and setting), participant characteristics (number, average age, gender, ethnicity, socioeconomic details and comorbidities), any reported subgroups of participants, tool used, if tool is validated, data collection method, outcome(s) measured (eg, repeat self-harm), method of outcome assessment, total number followed up, lost to follow-up number and reasons, number of events for each outcome and data on measures of association between assessment tool and outcome (eg, relative risk (RR), sensitivity and specificity) along with attendant precision (95% CI, p values) or raw data to calculate these. Unadjusted and adjusted data along with factors adjusted for were recorded. Where data were not reported, the corresponding author of the article was contacted by email. A follow-up email was sent if no response was received.

Quality assessment strategy

Risk of bias assessment of included studies used relevant elements of QUIPS[30 31] and QUADAS-2,[32] checklists suitable for prognostic and diagnostic tests, respectively. QUIPS covers areas of possible bias in the prognostic factor studies, such as study participation, attrition and outcome measurement, and QUADAS-2 covers areas of possible bias in diagnostic test accuracy studies, such as patient selection, index test and reference standard used. Studies reporting prognostic model development were assessed on a further four criteria identified from examining the quality assessment conducted for a systematic review of prognostic models.[33] These criteria include assessment of statistical methods used and methods of dealing with confounding factors.

Study selection, risk of bias assessment and data extraction were undertaken by two reviewers independently, and any disagreements resolved by discussion with referral to a third reviewer if required.

Methods of data analysis/synthesis

A narrative synthesis was carried out presenting information from all included studies to explore similarities, differences and findings within and between them. Where possible studies were grouped and analysed by tool.

Data synthesis

For each tool, outcome (self-harm; suicide) and outcome measure (OR, HR; sensitivity/specificity) grouping, contributing studies and data were assessed for clinical and methodological homogeneity. This informed a decision that meta-analysis was not appropriate, and therefore, data have been presented narratively.

Patient and public involvement

No patients or public were involved in this review.

RESULTS

Search results

A total of 17 137 potentially relevant records were identified through the literature search. After removal of duplicate records and screening for relevance to the review, inclusion criteria were applied to 103 full-text articles. Eleven studies[34–44] were included in the review and 92 excluded due to not meeting criteria for population (37 articles), study design (25), outcome (7), use of an assessment tool (15) and separately presenting data on an adolescent subgroup (8). Two of these last eight studies presented baseline, but not outcome, data for an adolescent subgroup so the authors were contacted to try to obtain these data. No response was received so these papers remain excluded (see figure 1).

Study characteristics

The 11 included studies comprised 8 prospective[35 36 38 40–44] and 3 retrospective[34 37 39] cohort designs and evaluated 10 tools (table 1); eight tools were specifically for self-harm/suicide assessment and two were for hopelessness and depression. Four studies attempted to develop a prediction model.[36 37 41 44] No trials evaluating the impact of using a tool were identified.

A total of 2554 participants were included, ranging from 10 to 24 years of age who were followed up for between 3 and 18 months. Three studies contained participants that had all presented with self-harm/suicide attempt,[34 36 43] and the remainder had mixed populations where >50% presented with self-harm/suicide attempt.[35 37–42 44] The proportion of each study population that presented with self-harm/suicide attempt is presented in table 1. As a result of the studies with mixed populations, 1818 of the total 2554 participants presented with self-harm/suicide.

The studies were carried out across a variety of settings in the UK[36 42] and USA.[34 35 37–41 43 44] Four took place in an emergency department,[34 37–39] four in inpatient units,[35 40 41 44] one each in an open treatment trial,[43] a home setting[36] and clinic sessions for a mixed inpatient/outpatient population.[42]

The 10 tools varied in length from three questions (Self-Assessed Expectation of Suicide Risk Scale)[37] to 93 questions (Self-Injurious Thoughts and Behaviours Interview, SITBI)[35]; the latter being unusually long as all other tools evaluated had ≤30 questions. The content assessed by the tools greatly varied. For example, eight of the tools asked about suicidal ideation, however, only two asked about previous self-harm. For full details of the
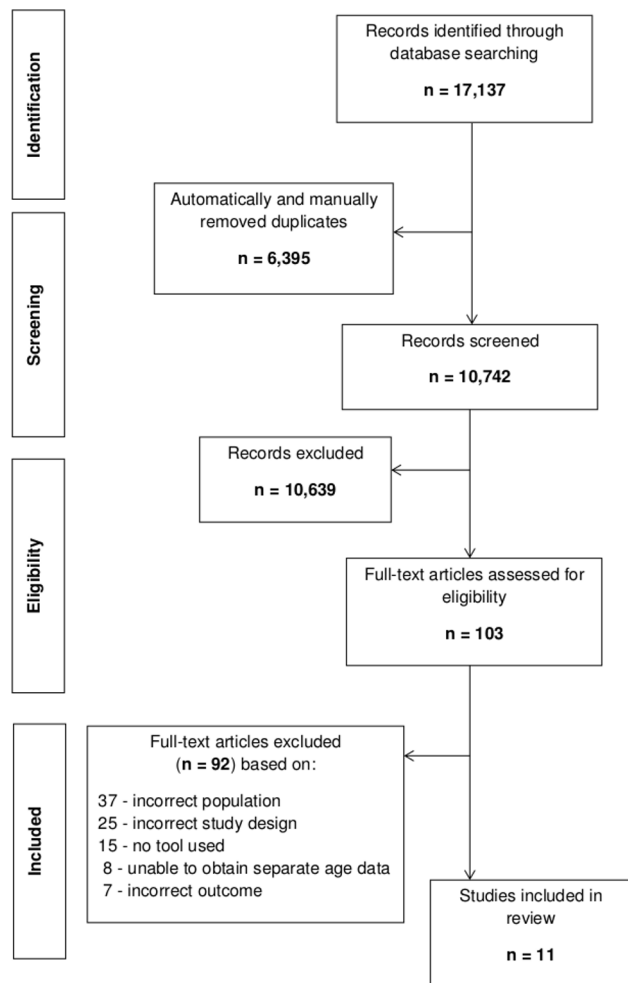
Figure 1 PRISMA flow diagram detailing the search process for included studies. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

tool content and length, please see the data presented in online supplementary appendix 2.

Across the studies, a selection of metrics to report outcome data were used. These include OR (adjusted and unadjusted), RR, HR (adjusted and unadjusted), and predictive validity statistics such as sensitivity and specificity. Online supplementary figure 1 details the full range of outcome measures, subgroups reported and factors adjusted for (if relevant) for each tool across all the studies.

Quality assessment

In general, reporting of the studies could have been better. For example, completeness of follow-up was rarely reported, and neither were details regarding whether the method of identifying the outcome (self harm/suicide attempt) was interpreted without knowledge risk assessment tool findings.

Only three of the studies[34 36 43] comprised a population where all patients presented with an episode of self-harm or a suicide attempt. In the remaining studies,[35 37–42 44] the proportion ranged from 51% to 79% but data were not separately reported for this group in these mixed population studies.

The four studies reporting prognostic models varied in quality.[36 37 41 44] All four were at the model development stage and only one of these was assessed as having a low risk of bias due to under reporting of details in the others.

Table 2 details the quality assessment undertaken on all 11 included studies.

FINDINGS

The findings from all included studies are presented by outcome measure (HR, OR and area under the curve (AUC) from receiver operator characteristic curves) in table 3 and predictive validity statistics (sensitivity, specificity, positive predictor value and negative predictor value) in table 4. Sensitivity can be defined as the proportion of positives that are correctly identified as positive, and specificity as the proportion of negatives that are correctly identified as negative. Positive predictive value (PPV) tells you the proportion of positive test results that truly are positive, and negative predictive value (NPV) tells you the proportion of test results that are truly negative. PPV and NPV, unlike sensitivity and specificity, are dependent on disease prevalence, so careful consideration of the disease prevalence must be made when applying PPV and NPV values to a different clinical population.[45]

Predicting future self-harm

Three studies, assessing four tools, contributed data for predicting future self-harm.

The Self-Injury Implicit Association Test (SI-IAT)[35] scores were not statistically significantly predictive of repeat self-harm at 3 months, (unadjusted OR 3.10, 95% CI 0·39 to 9.94, p≥0·05).

The SITBI[35] scores were found to be statistically significantly predictive of repeat self-harm at 3 months follow-up (adjusted OR 1.82, 95% CI 1.25 to 2.65, p=0.002). This study (evaluating both SI-IAT and SITBI) was assessed as having a high risk of bias, mostly due to poor reporting.

At 6 months follow-up, the accuracy of the Suicide Ideation Questionnaire (SIQ)[36] to classify a patient as high or low risk for self-harm repetition was: sensitivity 27.3%, specificity 99.2%, PPV 85.7% and NPV 85.6%. This tool performed poorly at identifying high-risk patients, but performed well at identifying low-risk patients.

The predictive validity of the Self-Harm Questionnaire (SHQ)[42] was reported at 3 months. This was: sensitivity 94.7%, specificity 34.6%, PPV 25.4% and NPV 96.6%. This tool performed well at identifying high-risk patients, but performed poorly at identifying low-risk patients.

PREDICTING FUTURE SUICIDE ATTEMPT

Eight studies reporting seven tools for predicting future suicide attempt.

The predictive ability of the Ask Suicide Screening Questions (ASQ)[34] was reported at 6 months as: sensitivity 95.8%, specificity 5.8%, PPV 16.8% and NPV 87.5%.

Table 1 Methodological characteristics and outcomes of included studies

Outcome repeat self-harm

Study ID	Study design	Tool/scale	Context	Sampling and population	No	% with self-harm or suicide attempt	Mean age in years (SD)	Outcome events	Outcome measure	Length of follow-up	Method of outcome identification
Cha et al[35]	Prospective cohort	SI-IAT, SITBI	Inpatient unit in Boston, USA	Adolescents recently admitted to a psychiatric inpatient unit (age 10–17)	123	55.3	14.8 (1.5)	Not stated	Adjusted OR	3 months	Not stated
Chitsabesan et al[36]	Prospective cohort	SIQ	Home setting in Manchester, UK	Consecutive referrals of overdoses to CAMHS teams (under 16)	162	100	14.5 (1.2)	23	Sensitivity and specificity	6 months	Not stated
Ougrin and Boege[42]	Prospective cohort	SHQ	Clinical sessions in South London and Maudsley NHS Foundation Trust for in/outpatients	In/outpatients having clinical sessions (age 12–17)	100	71	15.1 (1.7)	19	Sensitivity and specificity	3 months	Clinical record review, searching for self-harm

Outcome suicide attempt

Study ID	Study design	Tool/scale	Context	Sampling and population	No	% with self-harm or suicide attempt	Mean age in years (SD)	Outcome events	Outcome measure	Length of follow-up	Method of outcome identification
Ballard et al[34]	Retrospective cohort	ASQ	A&E department in Baltimore, USA	Consecutive patients presenting to A&E with suicide ideation or attempt (age 8–18)	230	100	14.0 (2.2)	24	Sensitivity and specificity	6 months	Hospital's electronic health record database, searched for suicide attempt
Study ID	Study design	Tool/Scale	Context	Sampling and population	No	% with self-harm or suicide attempt	Mean age in years (SD)	Outcome events	Outcome measure	Length of follow-up	Method of outcome identification
Czyz et al[37]	Retrospective cohort	SAESRS, C-SSRS	Psychiatric emergency department of Midwestern US university hospital	Patients seeking emergency psychiatric services (age 13–24)	340	59	17.6 (3.3)	39	Adjusted HR, AUC	Mean follow-up 486.7 days (range 12–18 months)	Hospital's electronic health record database, searched for suicide attempt
Gipson et al[38]	Prospective cohort	C-SSRS	Psychiatric emergency department of Midwestern US university hospital	Patients seeking emergency psychiatric services (age 13–17)	178	50.6	15.3 (1.3)	12	Unadjusted OR	12 months	Hospital's electronic health record database, searched for suicide attempt
Horowitz et al[39]	Retrospective cohort	C-SSRS	Psychiatric emergency department of Midwestern US university hospital	Patients seeking emergency psychiatric services (age 15–24)	473	51.6	19.3 (2.9)	147	Adjusted OR	18 months	Hospital's electronic health record database, searched for suicide attempt
King et al[40]	Prospective cohort	SIQ-JR	Inpatient unit in Michigan, USA	Patients hospitalised for acute suicide risk (age 13–17)	354	79	15.6 (1.3)	60	Relative risk	12 months	DISC-IV Mood disorders module of the diagnostic interview schedule for children

Continued

Table 1 Continued

Outcome repeat self-harm

King et al[41]	Prospective cohort	SIQ-JR, BHS, CDRS-R	Inpatient unit in Michigan, USA	Patients hospitalised for acute suicide risk (age 13–17)	352	76.1	15.6 (1.3)	63	Unadjusted HR, adjusted HR	12 months	DISC-IV Mood disorders module of the diagnostic interview schedule for children
Posner et al[43]	Prospective cohort	C-SSRS	Open treatment trial for adolescents suicide attempters in USA	Adolescents who had made a suicide attempt 45 days before enrolment (age 12–18)	124	100	15.8 (1.5)	24	Unadjusted OR	6 months	Columbia Suicide History Form
Yen et al[44]	Prospective cohort	SIQ	Inpatient unit in USA	Adolescents receiving inpatient treatment, admitted for elevated suicide risk	119	>62.2	15.3 (1.4)	37	Unadjusted HR, adjusted HR	6 months	Adolescent Longitudinal Interval Follow-Up Evaluation

A&E, Accident and Emergency; ASQ, Ask Suicide Screening Questions; AUC, area under the curve; BHS, Beck Hopelessness Scale; CDRS-R, Children's Depression Rating Scale-Revised; C-SSRS, Columbia-Suicide Severity Rating Scale;DISC-IV, Diagnostic Interview Schedule for Children; NHS, National Health Service; SAESRS, Self-Assessed Expectations of Suicide Risk Scale; SHQ, Self-Harm Questionnaire; SI-IAT, Self-Injury Implicit Association Test; SIQ, Suicide Ideation Questionnaire; SIQ-JR, Suicide Ideation Questionnaire-Junior; SITBI, Self-Injurious Thoughts and Behaviours Interview.

The SIQ-Junior (SIQ-JR) was evaluated in two studies, one reported RR[40] and the other reported adjusted and unadjusted HRs.[41] For every 1-point increase in SIQ-JR score, the RR of no future attempt was 0.93.[40] No CIs or p value were stated. For every 10-point increase in the SIQ-JR score, the unadjusted HR of future suicide attempt was 1.30 (95% CI 1.14 to 1.48, $p \leq 0.001$).[41] A subsequent multivariate regression model then reported an adjusted HR of 1.23 (95% CI 1.08 to 1.40, p=0.003).

The Columbia-Suicide Severity Rating Scale (C-SSRS) was evaluated by four studies,[37–39 43] reporting adjusted and unadjusted ORs, adjusted HR and AUC. Three of these studies reported outcomes for the C-SSRS, and the two subscales (severity and intensity) that the C-SSRS comprises.

The unadjusted OR ranged from 1.09 (95% CI 1.01 to 1.17) to 3.85 (95% CI 1.07 to 13.86) for every 1-point increase in C-SSRS score, and the adjusted OR ranged from 1.15 (95% CI 1.03 to 1.29) to 1.51 (95% CI 1.24 to 1.84) for every 1-point increase in C-SSRS score.

For the SIQ, unadjusted and adjusted HRs were reported.[44] In the univariate regression conducted, a statistically significant HR of 1.01 (95% CI 1.00 to 1.02, $p \leq 0.05$) for the dichotomised (high/low suicidal intent) SIQ score was reported, however, after the multivariate analysis, despite the HR being the same at 1.01 (95% CI 1.00 to 1.02, $p \geq 0.05$), it was no longer statistically significant.

For the Beck Hopelessness Scale,[41] an unadjusted HR of 1.51 (95% CI 1.22 to 1.87, $p \leq 0.001$) was reported for a 5-point increase in the scale.

For the Children's Depression Rating Scale-Revised[41] an unadjusted HR of 1.29 (95% CI 1.10 to 1.52, p=0.002) for a 10-point increase in the scale was reported.

DISCUSSION
Main findings
This systematic review identified 11 studies evaluating 10 tools for the prediction of future self-harm/suicide attempt in adolescent patients. Of these 11 studies, four attempted to develop a predictive model using the tools, while seven evaluated the tools as standalone instruments.

This review found there is a wide variation in the setting, population, outcome measure, length of follow-up and length and content of the tools in use. The variation was apparent in the predictive ability reported across the tools. The studies reporting predictive validity statistics varied greatly in results, for example, the SHQ had high sensitivity (94.7%) and low positive predictor value (25.4%),[42] whereas the SIQ had low sensitivity (27.3%) and high positive predictor value (85.7%).[36] It could be argued that a tool with higher sensitivity would be best suited for use in this population, to ensure that all those at risk are identified and offered treatment. However, this would cause issues with overtreatment and not be the best use of resources as the high sensitivity comes with low positive predictor values.

This finding of variation in predictive ability implies these tools are not sufficiently accurate to be used as standalone predictors of future risk of self-harm/suicide in adolescent populations, mirroring the findings of other reviews in adult populations.[3 25–27] This adolescent review and the other adult population reviews demonstrate there is insufficient evidence for use of such tools as the only component of a risk assessment, despite their widespread use for this purpose in clinical practice.[21]

Table 2 Quality assessment of included studies

Study ID	Ballard (2017)[34]	Cha (2016)[35]	Gipson (2015)[38]	Horowitz (2015)[39]	King (2013)[40]	Ougrin (2007)[42]	Posner (2013)[43]	Chitsabesan (2003)[36]	Czyz (2016)[37]	King (2010)[41]	Yen (2013)[44]
Was enrolment consecutive?	Yes	No	No	No	Unclear	Yes	Yes	Yes	Yes	Unclear	Unclear
What is the risk of selection bias?	Low risk	Moderate risk*	Moderate risk†	Moderate risk†	Low risk	Moderate risk‡	Low risk	Moderate risk§	Low risk	Low risk	Low risk
Index test results interpreted without knowledge of the results of the reference standard?	Unclear	Yes	Yes	Unclear	Yes	Yes	Yes	Yes	Unclear	Yes	Yes
Is the study population a mixed population in terms of presenting condition?	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Could the conduct or interpretation of the index test have introduced bias?	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Reference standard likely to correctly classify the target condition?	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes
Reference standard results interpreted without knowledge of the results of the index test?	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear	Yes	Unclear	Unclear	Unclear	Unclear

Continued

Table 2 Continued

Study ID	Ballard (2017)[34]	Cha (2016)[35]	Gipson (2015)[38]	Horowitz (2015)[39]	King (2013)[40]	Ougrin (2007)[42]	Posner (2013)[43]	Chitsabesan (2003)[36]	Czyz (2016)[37]	King (2010)[41]	Yen (2013)[44]
Could the conduct or interpretation of the reference standard have introduced bias?	Low risk	Unclear	Unclear	Low risk	Low risk	Low risk	Low risk	Unclear	Low risk	Low risk	Low risk
Appropriate interval between index test and reference standard?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
All patients received the same reference standard?	Yes	Unclear	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Yes	Yes
All patients included in the analysis?	Yes	No	Unclear	Yes	No	Yes	No	No	Yes	No	No
Was follow-up complete, and if not, were the reasons to lost to follow-up described and explored?	No and no	No and no	Unclear	Yes	No and yes	Yes	No and yes	No and no	Yes	No and yes	No and no
Overall rating of bias	Low	High	Unclear	Unclear	Low	Unclear	Low	Unclear	Unclear	Low	Unclear
Target condition/prognostic factor measured similarly to assign people to exposed and unexposed groups?								Yes	Yes	Yes	Yes
Confounding factors identified?								Yes	No	No	No

Continued

Table 2 Continued

Study ID	Ballard (2017)[34]	Cha (2016)[35]	Gipson (2015)[38]	Horowitz (2015)[39]	King (2013)[40]	Ougrin (2007)[42]	Posner (2013)[43]	Chitsabesan (2003)[36]	Czyz (2016)[37]	King (2010)[41]	Yen (2013)[44]
Strategies to deal with confounding factors stated?								Yes	No	No	No
Appropriate statistical analysis used?								Yes	Yes	Yes	Yes
Additional prognostic model criteria rating of bias								Low	Unclear	Unclear	Unclear

*Recruitment flow unclear. A total of 249 patients approached, but only 137 consented. No explanation of differences between those who did and did not consent.
†Only those residing in local county included.
‡Only patients from clinicians volunteering to take part in study were included.
§Inclusion only of drug overdoses.

Methodological limitations

Due to the high heterogeneity between included studies, specifically in terms of setting, population and length and content of tools used, no meta-analyses were conducted. Despite this, a wide range of data has been presented for the studies. Data were presented for each tool by outcome, and then compared by outcome measure reported where possible.

Also, use of the tools in differing populations and use of different outcome measures made comparing results impossible in studies reporting outcome measures such as HR and OR. For example, Horwitz et al[39] evaluated the severity subscale of the C-SSRS, as did Posner et al,[43] however, Horwitz used the subscale on a continuous basis, whereas Posner dichotomised the scores, meaning they were not directly comparable. The reporting of results using different measures is a known difficulty commonly encountered in systematic reviews evaluating diagnostic/predictive methods.[46] Very few studies reported all the metrics required for risk prediction.

Despite these limitations, the review has shown that multiple scales are in use for adolescent patients and that there is lack of consistency of practice. Additionally, the evidence from this review demonstrates that tools for adolescents have the same issues as tools for adults in terms of varying greatly in length and content.[22–24 47–50]

Implications of evidence

There are difficulties in drawing conclusions regarding the clinical significance of results and implications of the evidence from this review. First, the outcome data presented for each tool does not clearly translate into individual patient consequence or benefit.

Second, the very different settings the tools were evaluated in (eg, inpatient units, home treatment settings and emergency departments) means that results would not be transferable between these settings and the differing clinical populations within.

Third, some of the tools used and models developed are not practical for routine clinical use, particularly in an emergency department setting. For example, the model developed by Chitsabesan et al[36] requires four separate tools to be administered to both parent and child, and the subsequent scores to be used within a formula. This would clearly require refinement as it is currently a complex and time consuming process.

While the main aim of the review was to evaluate the ability of risk assessment tools to predict future episodes of suicide or self-harm, several other issues were identified that merit further investigation.

First, as seen in online supplemental appendix 2, the tools specific to self-harm/suicide vary greatly in content assessed. For example, seven out of these eight tools assess suicidal ideation, however, only two assess previous self-harm. As there is known to be a strong association between suicide and previous self-harm,[6 8 9] it seems surprising that only a minority of tools assess this important risk factor. Further primary research evaluating tool content and

Table 3	HR, OR, AUC results						
Outcome: Future suicide attempt							
Scale/tool	Subscale/subgroup (if applicable)	How score from scale/tool used	Outcome measure	Value	95% CI	P value	Study ID
BHS		Continuous measure (5-point increment)	Unadjusted HR	1.51	1.22 to 1.87	<0.001	King et al[41]
CDRS-R		Continuous measure (10-point increment)	Unadjusted HR	1.29	1.10 to 1.52	<0.001	King et al[41]
C-SSRS		Dichotomised (with/without ideation)	Adjusted HR	1.50	1.24 to 1.81	<0.001	Czyz et al[37]
		Dichotomised (with/without ideation)	AUC	0.74	n/a	<0.001	Czyz et al[37]
		Continuous measure (1 SD increment)	Unadjusted OR	1.45	1.07 to 1.98	0.02	Posner et al[43]
	Intensity subscale, ideators	Continuous measure (1-point increment)	Adjusted OR	1.15	1.03 to 1.29	<0.05	Horwitz (2015)[39]
		Continuous measure (1-point increment)	Unadjusted OR	1.10	0.97 to 1.28	>0.05	Gipson (2015)[38]
	Intensity subscale	Continuous measure (1-point increment)	Unadjusted OR	1.09	1.01 to 1.17	0.02	Gipson (2015)[38]
	Severity subscale	Continuous measure (1-point increment)	Adjusted OR	1.51	1.24 to 1.84	<0.001	Horwitz (2015)[39]
		Continuous measure (1 SD increment)	Unadjusted OR	1.43	0.99 to 2.05	0.05	Posner (2011)[43]
		Continuous measure (1-point increment)	Unadjusted OR	1.34	0.99 to 1.80	0.06	Gipson (2015)[38]
	Severity subscale, CSHF reference standard	Dichotomised (with/without intent)	Unadjusted OR	3.26	1.02 to 10.45	0.047	Posner (2011)[43]
	Severity subscale, SEB reference standard	Dichotomised (with/without intent)	Unadjusted OR	3.85	1.07 to 13.86	0.039	Posner (2011)[43]
SAESRS		Continuous measure	AUC	0.79	n/a	<0.001	Czyz (2016)[37]
Scale/tool	Subscale/subgroup (if applicable)	How score from scale/tool used	Outcome measure	Value	95% CI	P value	Study ID
SAESRS and C-SSRS		Continuous measure	Adjusted HR	0.94	0.90 to 0.97	0.001	Czyz (2016)[37]
		Continuous measure	AUC	0.8	n/a	<0.001	Czyz (2016)[37]

Table 3 Continued

Outcome: Future suicide attempt

SIQ		Dichotomised (high/low suicidal ideation)	Adjusted HR	1.01	1.00 to 1.02	>0.05 (NS)	Yen (2013)[44]
		Dichotomised (high/low suicidal ideation)	Unadjusted HR	1.01	1.00 to 1.02	<0.05	Yen (2013)[44]
SIQ-JR		Continuous measure (10-point increment)	Adjusted HR	1.23	1.08 to 1.40	0.003	King (2010)[41]
		Continuous measure (10-point increment)	Unadjusted HR	1.3	1.14 to 1.48	<0.001	King (2010)[41]
		Continuous measure (1-point increment)	Relative risk	0.93	Not stated	Not stated	King (2014)[40]

Outcome: Future self-harm

Scale/tool	Subscale/subgroup	How score from scale/tool used	Outcome measure	Value	95% CI	P value	Study ID
SI-IAT		Continuous measure	Adjusted OR	3.10	0.39 to 9.94	>0.05	Cha (2016)[35]
SITBI		Dichotomised (high/low likelihood of repetition)	Adjusted OR	1.82	1.25 to 2.65	0.002	Cha (2016)[35]

AUC, area under the curve; BHS, Beck Hopelessness Scale; CDRS-R, Children's Depression Rating Scale-Revised; CSHF, Columbia Suicide History Form; C-SSRS, Columbia-Suicide Severity Rating Scale; SAESRS, Self-Assessed Expectations of Suicide Risk Scale; SEB, Suicide Evaluation Board; SI-IAT, Self-Injury Implicit Association Test; SIQ, Suicide Ideation Questionnaire; SIQ-JR, Suicide Ideation Questionnaire-Junior; SITBI, Self-Injurious Thoughts and Behaviours Interview.

identifying predictive risk factors would be beneficial. Currently, available evidence is limited by dataset size and the populations included, for example, validation of tools on community populations rather than clinical ones.

Related to this is how the answers to the questions asked by the tools are used in predicative analysis. For example, the SHQ evaluated by Ougrin and Boege,[42] consists of three initial screening questions that everyone is asked establishing the presence of self-harm/suicidal ideation, and then a further 12 questions exploring self-harm for those who give a positive response to the initial three. The analysis conducted by Ougrin and Boege, however, only used the first three questions to classify patients into positive/negative, and the second section of 12 questions was not used. As a result, the basis of classification in this study was effectively history of self-harm/suicidal ideation, rather than the use of the tool's questions about self-harm to stratify risk and predict which patients would go on to repeat self-harm. Further analysis of this tool with regard to the questions asked in each section would be beneficial to explore the predicative ability of this additional data collected.

Second, no distinction was found about whether the tools in use were for patients presenting with self-harm/suicide attempt for the first time, or those presenting with repeat events. As some tools asked about previous self-harm/suicide attempts, this was perhaps being taken into account, but research exploring whether these patient groups have the same risk factors and how previous events can modify them, would be useful to understand further how to predict and prevent this behaviour, and determine whether separate assessments would be appropriate.

Third, Czyz et al[37] explored self-rated suicide risk, as opposed to that rated through tools. They found that asking adolescents to self-rate their risk of suicide was more accurately predictive of future suicide attempts, and this was of particular relevance in busy emergency departments with time pressures. This was the only study exploring self-rated risks, however, and more primary studies are required to explore further the use of self-rating in predicting future risk. This could have important implications for changing practice with regard to how suicide risk is assessed.

Fourth, this review attempted to explore the use of prognostic models in self-harm/suicide prediction, however,

Table 4 Predictive validity results

Outcome	Scale/tool	How score from scale/tool used	True positives	False positives	True negatives	False negatives	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictor value	Negative predictor value	Study ID
Future suicide attempt	ASQ	Dichotomised (positive/negative suicide risk)	23	114	7	1	95.8 (78.8 to 99.9)	5.8 (2.4 to 11.6)	16.8 (15.5–18.1)	87.5 (47.4–98.1)	Ballard et al[34]
Future self-harm	SHQ	Dichotomised (SHQ positive/negative)	18	53	28	1	94.7 (74.0 to 99.9)	34.6 (24.3 to 46.0)	25.4 (21.9–29.1)	96.6 (80.2–99.5)	Ougrin et al[42]
	SIQ	Dichotomised (suicidal ideation yes/no)	6	1	124	16	27.3 (10.7 to 50.2)	99.2 (95.6 to 99.9)	85.7 (43.1–97.9)	88.5 (88.5–90.9)	Chitsabesant et al[36]

ASQ, Ask Suicide Screening Questions; SHQ, Self-Harm Questionnaire; SIQ, Suicide Ideation Questionnaire.

due to poor reporting and models only being in the early stages of development, this was not really possible in any great depth. The four studies that reported model development were difficult to understand from a modelling point of view. Methods were not explicitly stated making it hard to comprehend how analyses had been conducted, confounding factors were either not identified or methods for addressing them were not stated, and lack of full reporting meant that carrying out a detailed quality appraisal was not possible. Future work should build on these early modelling studies to further develop them to the later stages of internal and external validation and impact studies, and it is imperative that future studies should be reported fully.

CONCLUSION

This review is the first to explore the use of tools to predict future self-harm/suicide attempts in an adolescent population. It has shown that the current limited amount of primary evidence means at present no individual tool can be identified as performing better than another or is a sufficiently accurate predictor of self-harm/suicide risk. The use of tools in prediction models is at an early stage and further research is warranted to develop these further. Risk assessment tools should only be used as part of a wider comprehensive assessment, assessing more than just the risk of repetition. This reflects current standard practice in the UK, where National Institute for Health and Care Excellence guidelines recommend thorough psychosocial needs assessment focused on patient needs rather than risk.[18–20]

REFERENCES

1. Hawton K, Saunders KEA, O'Connor RC. Self-Harm and suicide in adolescents. *The Lancet* 2012;379:2373–82.
2. Hawton K, Bergen H, Waters K, *et al*. Epidemiology and nature of self-harm in children and adolescents: findings from the multicentre study of self-harm in England. *Eur Child Adolesc Psychiatry* 2012;21:369–77.

3. Chan MKY, Bhatti H, Meader N, *et al*. Predicting suicide following self-harm: systematic review of risk factors and risk scales. *Br J Psychiatry* 2016;209:277–83.

4. Madge N, Hewitt A, Hawton K, *et al*. Deliberate self-harm within an international community sample of young people: comparative findings from the Child & Adolescent Self-harm in Europe (CASE) Study. *J Child Psychol & Psychiat* 2008;49:667–77.

5. De Leo D, Heller TS. Who are the kids who self-harm? an Australian self-report school survey. *Med J Aust* 2004;181:140–4.

6. Hawton K, Bergen H, Kapur N, *et al*. Repetition of self-harm and suicide following self-harm in children and adolescents: findings from the multicentre study of self-harm in England. *J Child Psychol Psychiatry* 2012;53:1212–9.

7. Hawton K, Harriss L. Deliberate self-harm by under-15-year-olds: characteristics, trends and outcome. *J Child Psychol Psychiatry* 2008;49:441–8.

8. Hawton K, Bergen H, Cooper J, *et al*. Suicide following self-harm: findings from the multicentre study of self-harm in England, 2000-2012. *J Affect Disord* 2015;175:147–51.

9. Suicide by Children and Young People in England. *National confidential inquiry into suicide and homicide by people with mental illness (Ncish)*. Manchester: University of Manchester, 2016.

10. Hawton K, Hall S, Simkin S, *et al*. Deliberate self-harm in adolescents: a study of characteristics and trends in Oxford, 1990-2000. *J Child Psychol Psychiatry* 2003;44:1191–8.

11. Cooper J, Kapur N, Webb R, *et al*. Suicide after deliberate self-harm: a 4-year cohort study. *Am J Psychiatry* 2005;162:297–303.

12. Guan K, Fox KR, Prinstein MJ. Nonsuicidal self-injury as a time-invariant predictor of adolescent suicide ideation and attempts in a diverse community sample. *J Consult Clin Psychol* 2012;80:842–9.

13. Muehlenkamp JJ. Self-injurious behavior as a separate clinical syndrome. *Am J Orthopsychiatry* 2005;75:324–33.

14. Madge N, Hawton K, McMahon EM, *et al*. Psychological characteristics, stressful life events and deliberate self-harm: findings from the Child & Adolescent Self-harm in Europe (CASE) Study. *Eur Child Adolesc Psychiatry* 2011;20:499–508.

15. Brophy M. *Truth hurts: report of the National inquiry into self-harm among young people*, 2006.

16. Evans E, Hawton K, Rodham K. Factors associated with suicidal phenomena in adolescents: a systematic review of population-based studies. *Clin Psychol Rev* 2004;24:957–79.

17. Abraham MB, Nicholas JA, Smith GJ, *et al*. Reduction in hypoglycemia with the predictive low-glucose management system: a long-term randomized controlled trial in adolescents with type 1 diabetes. *Diabetes Care* 2018;41:303–10.

18. National Collaborating Centre for Mental Health. *Self-Harm: longer term management NICE clinical guideline 133*. London: National Institute for Clinical Excellence, 2011.

19. National Collaborating Centre for Mental Health. *Self-Harm: the short-term physical and psychological management and secondary prevention of self-harm in primary and secondary care NICE clinical guideline 16*. London: National Institute for Clinical Excellence, 2004.

20. National Institute for Health and Care Excellence. *Self–Harm. NICE quality standard QS34*. London: NICE, 2013.

21. Quinlivan L, Cooper J, Steeg S, *et al*. Scales for predicting risk following self-harm: an observational study in 32 hospitals in England. *BMJ Open* 2014;4:e004732.

22. Phull JS, Singh Phull J. A survey of assessment tools for detecting and managing risk. *Mental Health Practice* 2012;15:31–3.

23. Hawley CJ, Littlechild B, Sivakumaran T, *et al*. Structure and content of risk assessment proformas in mental healthcare. *J Ment Health* 2006;15:437–48.

24. Higgins N, Watts D, Bindman J, *et al*. Assessing violence risk in general adult psychiatry. *Psychiatr Bull* 2005;29:131–3.

25. Quinlivan L, Cooper J, Davies L, *et al*. Which are the most useful scales for predicting repeat self-harm? A systematic review evaluating risk scales using measures of diagnostic accuracy. *BMJ Open* 2016;6:e009297.

26. Carter G, Milner A, McGill K, *et al*. Predicting suicidal behaviours using clinical instruments: systematic review and meta-analysis of positive predictive values for risk scales. *Br J Psychiatry* 2017;210:387–95.

27. Randall JR, Colman I, Rowe BH. A systematic review of psychometric assessment of self-harm risk in the emergency department. *J Affect Disord* 2011;134:348–55.

28. Açbay O, Gündoğdu S. Can metformin reduce insulin resistance in polycystic ovary syndrome? *Fertil Steril* 1996;65:946–9.

29. Kokkevi A, Rotsika V, Arapaki A, *et al*. Adolescents' self-reported suicide attempts, self-harm thoughts and their correlates across 17 European countries. *J Child Psychol Psychiatry* 2012;53:381–9.

30. Hayden JA, Côté P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med* 2006;144:427–37.

31. Hayden JA, van der Windt DA, Cartwright JL, *et al*. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158:280–6.

32. Whiting PF, Rutjes AWS, Westwood ME, *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.

33. Elias A, Mallett S, Daoud-Elias M, *et al*. Prognostic models in acute pulmonary embolism: a systematic review and meta-analysis. *BMJ Open* 2016;6:e010324.

34. Ballard ED, Cwik M, Van Eck K, *et al*. Identification of at-risk youth by suicide screening in a pediatric emergency department. *Prev Sci* 2017;18:174–82.

35. Cha CB, Augenstein TM, Frost KH, *et al*. Using implicit and explicit measures to predict Nonsuicidal self-injury among adolescent inpatients. *J Am Acad Child Adolesc Psychiatry* 2016;55:62–8.

36. Chitsabesan P, Harrington R, Harrington V, *et al*. Predicting repeat self-harm in children--how accurate can we expect to be? *Eur Child Adolesc Psychiatry* 2003;12:23–9.

37. Czyz EK, Horwitz AG, King CA. Self-Rated expectations of suicidal behavior predict future suicide attempts among adolescent and young adult psychiatric emergency patients. *Depress Anxiety* 2016;33:512–9.

38. Gipson PY, Agarwala P, Opperman KJ, *et al*. Columbia-suicide severity rating scale: predictive validity with adolescent psychiatric emergency patients. *Pediatr Emerg Care* 2015;31:88–94.

39. Horwitz AG, Czyz EK, King CA. Predicting future suicide attempts among adolescent and emerging adult psychiatric emergency patients. *J Clin Child Adolesc Psychol* 2015;44:751–61.

40. King CA, Jiang Q, Czyz EK, *et al*. Suicidal ideation of psychiatrically hospitalized adolescents has one-year predictive validity for suicide attempts in girls only. *J Abnorm Child Psychol* 2014;42:467–77.

41. King CA, Kerr DCR, Passarelli MN, *et al*. One-Year follow-up of suicidal adolescents: parental history of mental health problems and time to post-hospitalization attempt. *J Youth Adolesc* 2010;39:219–32.

42. Ougrin D, Boege I. Brief report: the self harm questionnaire: a new tool designed to improve identification of self harm in adolescents. *J Adolesc* 2013;36:221–5.

43. 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;168:1266–77.

44. Yen S, Weinstock LM, Andover MS, *et al*. Prospective predictors of adolescent suicidality: 6-month post-hospitalization follow-up. *Psychol Med* 2013;43:983–93.

45. Attia J. Diagnostic tests: moving beyond sensitivity and specificity: using likelihood ratios to help interpret diagnostic tests. *Aust Prescr* 2003;26:111–3.

46. Walter SD, Sinuff T. Studies reporting ROC curves of diagnostic and prediction data can be incorporated into meta-analyses using corresponding odds ratios. *J Clin Epidemiol* 2007;60:530–4.

47. Cooper J, Kapur N, Dunning J, *et al*. A clinical tool for assessing risk after self-harm. *Ann Emerg Med* 2006;48:459–66.

48. Patterson WM, Dohn HH, Bird J, *et al*. Evaluation of suicidal patients: the sad persons scale. *Psychosomatics* 1983;24:343–9.

49. Spittal MJ, Pirkis J, Miller M, *et al*. The repeated episodes of self-harm (RESH) score: a tool for predicting risk of future episodes of self-harm by hospital patients. *J Affect Disord* 2014;161:36–42.

50. Steeg S, Kapur N, Webb R, *et al*. The development of a population-level clinical screening tool for self-harm repetition and suicide: the react self-harm rule. *Psychol Med* 2012;42:2383–94.