Predictive accuracy of risk scales following self-harm: multicentre, prospective cohort study

Leah Quinlivan, Jayne Cooper, Declan Meehan, Damien Longson, John Potokar, Tom Hulme, Jennifer Marsden, Fiona Brand, Kezia Lange, Elena Riseborough, Lisa Page, Chris Metcalfe, Linda Davies, Rory O'Connor, Keith Hawton, David Gunnell and Nav Kapur

Background

Scales are widely used in psychiatric assessments following self-harm. Robust evidence for their diagnostic use is lacking.

Aims

To evaluate the performance of risk scales (Manchester Self-Harm Rule, ReACT Self-Harm Rule, SAD PERSONS scale, Modified SAD PERSONS scale, Barratt Impulsiveness Scale); and patient and clinician estimates of risk in identifying patients who repeat self-harm within 6 months.

Method

A multisite prospective cohort study was conducted of adults aged 18 years and over referred to liaison psychiatry services following self-harm. Scale *a priori* cut-offs were evaluated using diagnostic accuracy statistics. The area under the curve (AUC) was used to determine optimal cut-offs and compare global accuracy.

Results

In total, 483 episodes of self-harm were included in the study. The episode-based 6-month repetition rate was 30% ($n = 145$). Sensitivity ranged from 1% (95% CI 0–5) for the SAD PERSONS scale, to 97% (95% CI 93–99) for the Manchester Self-Harm Rule. Positive predictive values ranged from 13% (95% CI 2–47) for the Modified SAD PERSONS Scale to 47% (95% CI 41–53) for the clinician assessment of risk. The AUC ranged from 0.55 (95% CI 0.50–0.61) for the SAD PERSONS scale to 0.74 (95% CI 0.69–0.79) for the clinician global scale. The remaining scales performed significantly worse than clinician and patient estimates of risk ($P < 0.001$).

Conclusions

Risk scales following self-harm have limited clinical utility and may waste valuable resources. Most scales performed no better than clinician or patient ratings of risk. Some performed considerably worse. Positive predictive values were modest. In line with national guidelines, risk scales should not be used to determine patient management or predict self-harm.

Self-harm presentations to hospital emergency departments are common and have serious consequences.[1,2] In the UK, the risk of suicide has been reported to be approximately 50 times greater for patients in the year after a self-harm episode compared with the general population.[3] It may be even higher for those with repeated episodes.[4,5] Good-quality assessment of people when they present to hospital with self-harm is a core part of clinical practice in many countries and can reduce risk of repeat self-harm.[6–8] Following an initial assessment by emergency department staff, liaison psychiatry clinicians may subsequently provide a more comprehensive evaluation of needs and risk, often including formal risk scales.[6,9,10] The use of risk scales in the assessment of self-harm is contentious,[6,11,12] with some clinical guidelines advocating the use of psychometrically tested scales over locally developed proformas[11] and others suggesting that risk instruments should not be used to predict outcome but may be used to help structure assessments.[6] Despite limited evidence for their effectiveness, risk scales are in widespread use in hospital services. Our recent study of 32 hospitals in England found over 20 tools in use, indicating uncertainty over which are the best scales following self-harm.[12] This reflects the situation internationally.[8] This uncertainty concerning risk-prediction scales may be indicative of the inconsistency in the evidence base.[6,13]

Our recent systematic review of cohort studies evaluating the predictive accuracy of scales included 8 studies[14–21] and 11 scales. Sensitivity for identifying repeat episodes ranged from 97% for the Manchester Self-Harm Rule[17] to 3% for the Modified SAD PERSONS scale,[22] and positive predictive values ranged from 70% for the Barratt Impulsiveness Scale[23] to 7% for the Modified SAD PERSONS Scale.[13] Other reviews report similarly variable performance across risk scales.[8,24–26]

On the basis of the published work it is not possible to identify the best performing scale. Direct comparison of different instruments between studies is not appropriate because of wide variations in methodological quality, case mix, study setting, scoring thresholds, follow-up and reporting. Analyses also tend to be restricted to those based on available contingency tables (for example, sensitivity, specificity, positive and negative predictive values). Without access to raw data, it is not possible to investigate more comprehensive measures of performance such as the 'area under the receiver operating characteristic curve', which evaluates the performance of a scale at different thresholds.[13,27] In order to compare different risk scales, we tested the predictive utility of widely used instruments as well as clinician and patient-rated global measures of risk in a multicentre cohort study in England with a 6-month follow-up. Our overall aim was

to compare the performance of the scales in people who were referred from the emergency department to psychiatric liaison services following self-harm. This evidence is important for clinicians, service providers, commissioners and policymakers in order to critically evaluate the use of risk scales. To increase the ecological validity of the study, the clinicians administered the risk assessments as close to the usual psychiatric assessment as possible. Our specific objectives were to: (a) estimate the predictive accuracy of the scales for repeat self-harm using published cut-offs; (b) plot receiver operating characteristic (ROC) curves and examine the area under the curve (AUC) for each of the scales; (c) estimate the predictive accuracy of the scales for repeat self-harm using data-determined optimal cut-offs that maximise sensitivity and specificity in this sample. We hypothesised that specific scales, which are often based on the most important epidemiological risk factors,[6,13–21] would perform better than global measures of clinician- or patient-rated risk.

Method

We conducted a multicentre prospective cohort study to examine the diagnostic accuracy of risk scales for repeat self-harm. The study was reviewed and approved by the Central Manchester Research Ethics Committee (REC No: 13/NW/0838) prior to commencement.

Inclusion and exclusion criteria

Participants were patients aged 18 years or over who were referred from emergency departments to psychiatric liaison services for assessment following self-harm in five large teaching hospitals in England (Brighton, Bristol, Derby, Manchester and Oxford), between March 2014 and January 2015. We did not include people under the age of 18 years because, in the UK, service provision is different for younger people.[28,29] We focused on people who received a psychiatric assessment because risk assessments are a key component of psychiatric practice and are in widespread use.[6,12] People who are referred for psychiatric assessment may also be at higher risk of adverse outcomes than patients who are not referred.[30,31] Previous research suggests that patients who present with self-harm and receive a psychiatric assessment are older, and less likely to be unemployed or use self-cutting as a method of self-harm and more likely to have factors suggestive of current mental illness or treatment than those who were not assessed.[32] People who were unable to understand English were excluded as the risk scales have not been translated or tested in non-English-speaking groups. People who were unable to consent (for example, because of impaired consciousness or active psychosis) or who were deemed too unwell or aggressive to participate by the clinical team were also excluded. Episodes where the patient did not stay for psychiatric assessment or treatment were also not included.

Service provision

The five research sites were based in urban areas and varied in population size (150 000–500 000), deprivation (from the 5th most deprived area out of 326 local authorities to the 166th), ethnicity (proportion of individuals from Black and minority ethnic groups 5.8–33.3%), and rates of unemployment (proportion unemployed 3.6–8.1%). The services for people who self-harm were provided by multidisciplinary psychiatric liaison teams that included psychiatric nurses, social workers, consultant psychiatrists and junior doctors. Junior doctors and/or crisis teams provided out-of-hours services at all sites. The

teams varied in their hours of operation; three were available 24 h, 7 days a week, and two of the teams were available from around 07.00 to 21.00. The proportion of patients who received a psychiatric assessment ranged from 45 to 77%. Consistent with National Institute for Health and Care Excellence (NICE) guidelines, each patient episode of self-harm is treated in its own right as a one-time visit.[6] The duration of assessments is dependent on patient need but they usually have a modal duration of around 1 h.

Case definition

There are debates over nomenclature in suicide prevention research and several terms are used to denote self-harm and suicidal behaviour.[33,34] Terms such as 'non-suicidal self-injury' (self-injury without intent) or suicide attempts (self-harm with suicidal intent) are frequently used to classify patients;[35] but focusing on specific methods and/or suicidal intent may be clinically problematic.[28,36] Suicidal behaviour is often characterised by ambivalence and changeability, intent may vary both between and within episodes, and even apparently low-intent episodes are associated with high mortality risk.[28,36,37] Therefore, consistent with national UK guidance[6] we included all presentations for self-harm in this study – defined as episodes of intentional self-injury or self-poisoning, irrespective of motivation or degree of suicidal intent.[38] The same definition was used across all research sites.

We calculated that a sample size of 480 would provide adequate statistical power to estimate diagnostic properties with reasonable precision (for example, assuming a repeat rate of 15%, the 95% confidence interval around a sensitivity of 0.80 would be 0.69–0.89) and also to detect a difference between the accuracy of scales. We therefore had a target sample size of approximately 100 per centre.[39,40]

Procedure

The research team met with clinicians at all sites on several occasions to familiarise them with the study procedure and risk scales and to answer any queries. In all cases clinicians (largely nurses and psychiatrists) obtained informed consent from patients as well as conducting the assessments. The assessments generally took place in the emergency department or on a medical ward. We adopted an episode-based approach to analysis, that is, we investigated repetition subsequent to each episode of self-harm, which meant that some individuals were included more than once. This more readily reflects the clinical reality of presentation to services[19,41] and is consistent with national guidelines that suggest each episode should be assessed in its own right.[6]

Scales

The assessment scales were selected for inclusion in the study on the basis of a systematic review of the diagnostic accuracy of risk scales in previous studies[13] as well as practical service considerations (such as time taken to complete the scale – a scale with a large number of items would be highly unlikely to be adopted in routine practice). They included basic clinical and demographic information. The Manchester Self-Harm Rule, ReACT Self-Harm Rule, Modified SAD PERSONS scale, and SAD PERSONS scale comprised items collected as part of the clinical assessment. The Barratt Impulsiveness Scale included specifically collected data. Because of this, it was not possible to randomise the order of administration of the scales. On the advice of the clinical teams and in order not to disrupt the routine

clinical assessments the study was in general introduced after the clinical interviews.

For the risk scales (Barratt Impulsiveness Scale,[23] Manchester Self-Harm Rule,[17] ReACT Self-Harm Rule,[19] SAD PERSONS scale,[42] Modified SAD PERSONS scale[22]) *a priori* cut-offs were chosen on the basis of previous literature.[13] It was, of course, not possible to have full masking of the scales, but clinicians were masked to how they were scored and the scoring thresholds. We also included a clinician and patient global evaluation of risk scale. These each consisted of a single question, which asked the respondent to estimate the likelihood of repeat self-harm within 6 months on a 1–10 Likert-type scale (for example: 'How likely do you think it is, that [you]/[the patient] will repeat self-harm within the next six months? Please indicate on this scale (with 1 as extremely unlikely and 10 extremely likely)'. We used the mid-point as our cut-off point (i.e. 0 to 5, low risk; 6+ high risk). Further details are available from the authors on request. Further details on all the scales, the assessment, and scoring are presented in online supplement DS1.

Reference standard

The outcome for the study was hospital-treated repeat self-harm within 6 months of presentation and was ascertained masked to index test results from hospital databases by linking National Health Service (NHS) and local hospital numbers where available. Where these were not available, cases were linked by using a combination of date of presentation, name and age. Teams in the individual centres carried out all data linkage and identifiable data were not passed to the research team. The date of each subsequent episode of self-harm was ascertained and clinicians used the standard definition of self-harm described above. The time frame of 6 months was selected, as this is a high-risk period and one that has often been used in previous studies.[43,44]

Analysis

Predictive accuracy statistics

The diagnostic accuracy of each of the scales in Table 1 was evaluated using a range of diagnostic accuracy statistics and 95% confidence intervals, including: sensitivity, specificity, negative/positive predictive values, positive/negative likelihood ratios and the diagnostic odds ratio using predetermined published cut-off points where available (see online supplement DS2 for definitions).[13] Meta-analysis using random-effects modelling (DerSimonian–Laird method)[45] was used to explore variation by centre in sensitivity and specificity. Heterogeneity was evaluated as present if Cochran's Q was less than 0.10 and Higgins I^2 was greater than 50%.[46] The scales were analysed separately. ROC curves, which plot sensitivity on the y-axis and $1-$specificity on the x-axis for all possible cut-off points, were constructed for each total scale score and overall discriminative ability was evaluated by the AUC.[27] Higher values for the AUC indicate greater discriminatory power. An AUC of 1.0 indicates a perfect test and 0.5 indicates the result is no better than chance.[47] We compared the formal scales to the clinician- and patient-rated global measures of risk by calculating the difference between the respective AUCs.[48] For our third aim, optimal cut-off points for our sample were selected using Youden's J index, which maximises the difference between true and false positive rates (provides the point with the furthest distance from the diagonal line).[27] Standard errors and exact binomial exact confidence intervals were calculated using the DeLong method.[48]

Scales were generally very well completed, with the exception of the 30-item Barratt Impulsiveness Scale, and we used

multiple regression imputation for episodes that had less than 5% missing data on this scale.[49] SPSS version 20, Stata 13.0 and OpenEpi (Open Source Epidemiologic Statistics for Public Health, www.OpenEpi.com) were used for the analyses.

Patient involvement

An expert-by-experience was a co-applicant on the NIHR Programme Grant and actively contributed to the study design. Patient advisors, carers and clinicians contributed to the research questions, scales (for example, we included patient and clinician global estimations of risk) and outcomes. There was also patient input into our dissemination plan, which includes dissemination to participants and the relevant patient community.

Results

Demographic characteristics

Participating clinicians considered 1301 patients referred to liaison psychiatry services after self-harm for inclusion in the study, of whom 421 were judged not to be appropriate (for example, they were too unwell, too distressed, intoxicated or in police custody) and 353 refused to participate. The consenting sample resulted in data on 514 separate episodes of self-harm that was reduced to 483 after exclusion of episodes with significant missing data on key scales. The 483 episodes of self-harm represented data from 464 separate individuals, with 12 individuals appearing in the data-set more than once. Psychiatric liaison nurses conducted the majority of the assessments ($n = 374/483$, 77.4%). Psychiatrists conducted 82/483 of the assessments (17%) and the remainder were conducted by junior doctors/nurses or other allied health professionals (such as social workers, therapists) that were attached to the clinical teams ($n = 26/483$, 5.4%). The assessor was unknown for one of the assessments (0.2%).

The median age for the sample was 33 years (interquartile range 22–42 years, range 18–88 years), and over half were women (298/483, proportion women 61.7%) and under 35 years of age (297/481, 61.7%). The majority of the sample was of White ethnicity (455/483, 94.2%). Many participants had a self-reported history of previous self-harm (359/483, 74.3%), and 245/483 (50.7%) had had an episode of self-harm in the past 12 months. Over half of the sample had a prior psychiatric history (310/483, 64.2%). The most common method of self-harm was self-poisoning (393/483, 81.4%), followed by self-cutting (71/483, 14.7%) and other methods (19/483, 3.9%) (for example drowning, asphyxiation). In the 6 hours prior to the self-harm episode just over half of the participants had used alcohol (248/470, 52.8%) and 11% recreational drugs (51/463). The episode-based 6-month self-harm repetition rate was 30% (145/483).

Scores on the risk scales and repeat self-harm

The distribution of the tested scale results using established cut-offs and for the clinician and patient global estimations of risk using the median cut-off points are presented in Table 1.

Performance of the scales

Diagnostic performance varied greatly. Sensitivity, which is the proportion of patients who repeated self-harm and were correctly identified by the scale as high risk, ranged from 1% for the SAD PERSONS scale using the moderate/high-risk threshold to 97% for the Manchester Self-Harm Rule and ReACT Rule at the recommended cut-off of one. Positive predictive values, which are the probability that a patient identified as high risk by the

Table 1 The distribution of the seven scales' results and repeat self-harm by 6 months according to predefined cut-off points

Scale, thresholds	Did not repeat ($n = 338$, 70%)	Repeat self-harm ($n = 145$, 30%)	Total ($n = 483$)
Manchester Self-Harm Rule			
Low risk (0)	67 (94.4)	4 (5.6)	71 (14.7)
Moderate/high risk (1+)	271 (65.8)	141 (34.2)	412 (85.3)
ReACT Self-Harm Rule			
Low risk (0)	79 (94.0)	5 (6.0)	84 (17.4)
Moderate/high risk (1+)	259 (64.9)	140 (35.1)	399 (82.6)
SAD PERSONS scale			
Low (0–4)	303 (71.3)	122 (28.7)	425 (88.0)
Moderate (5–6)	29 (58.0)	21 (42.0)	50 (10.4)
High (7–10)	6 (75.0)	2 (25.0)	8 (1.7)
Modified SAD PERSONS scale			
Low (0–5)	267 (72.0)	104 (28.0)	371 (76.8)
Moderate (6–8)	64 (61.5)	40 (38.5)	104 (21.5)
High (>8)	7 (87.5)	1 (12.5)	8 (1.7)
Clinician global scale			
<5	217 (85.1)	38 (14.9)	255 (52.8)
6+	121 (53.1)	107 (46.9)	228 (47.2)
Patient global scale			
<5	213 (82.9)	44 (17.1)	257 (53.2)
6+	125 (55.3)	101 (44.7)	226 (46.8)
Barratt Impulsiveness Scale[a]			
<96	331 (70.3)	140 (29.7)	471 (97.5)
97+	7 (58.3)	5 (41.7)	12 (2.5)

a. Cut-off based on Randall et al.[20]

test actually went on to repeat self-harm, were low for the high sensitivity scales (Manchester Self-Harm Rule and the ReACT Rule) (34 and 35%). Positive predictive values were highest for the clinician global estimation of risk scale (47%), followed by the patient global estimation of risk scale (44%), using the mid-point cut-off for each. The full range of diagnostic accuracy statistics for the scales when using *a priori* cut-off points are presented in Table 2, and those for optimised cut-offs, which maximise sensitivity and specificity according to Youden's *J* index, in Table 3.

Heterogeneity between sites

The performance of the Manchester Self-Harm Rule, ReACT Self-Harm Rule, the patient global scale, the Barratt Impulsiveness Scale and the Modified SAD PERSONS scale using *a priori* cut-offs was similar across sites, with no statistical evidence of heterogeneity. For the clinician global rating of risk, specificity varied from 58 to 82% ($\tau^2 = 0.16$, $\chi^2 = 13.4$, d.f. = 5, $P = 0.01$, $I^2 = 70.4\%$). Specificity also varied for the SAD PERSONS scale (between 74 and 96%), ($\tau^2 = 0.46$, $\chi^2 = 14.6$, d.f. = 5, $P < 0.01$, $I^2 = 73\%$).

Area under the receiver operating characteristic curves

The ROC curves that show the relationship between the sensitivity and specificity for the respective scales are presented in Fig. 1(a) and 1(b) along with the AUC and 95% confidence intervals for the respective scales. The AUCs of the seven scales varied between bordering on no better than chance for the SAD PERSONS scale (0.55) and Modified SAD PERSONS scale (0.58), to poor for the Barratt Impulsiveness Scale (0.62) and fair accuracy for the ReACT Self-Harm Rule (0.70), patients' estimation of risk (0.72), Manchester Self-Harm Rule (0.72) and clinicians' global estimation of risk (0.74).

Regarding global estimations of risk, both the clinician and patient scales were significantly better than the SAD PERSONS scale (AUC difference: 0.19, $P < 0.001$; AUC difference: 0.16, $P < 0.001$, respectively), the Modified SAD PERSONS scale (AUC difference: 0.16, $P < 0.001$; AUC difference: 0.13, $P < 0.001$) and the Barratt Impulsiveness Scale (AUC difference: 0.12, $P < 0.001$; AUC difference: 0.09, $P < 0.001$). There were no significant differences between the global estimates of risk and the remainder of the scales.

Discussion

Main findings

The results of this multisite prospective cohort study indicate that risk scales generally performed poorly in terms of predicting repeat self-harm. High sensitivity scales tended to have poor specificity and vice versa. Possible exceptions to this were clinician- and patient-rated measures of global risk. Contrary to our hypothesis, formal risk scales performed no better than the global assessments and in some cases there was evidence (on the basis of ROC curves) that performance was significantly worse. Using our available study data to select optimal cut-offs in this sample resulted in better performance (Table 3), but these were essentially *post hoc* estimations of 'best case' predictive utility and would not be generalisable to other samples. Our findings suggest that risk assessment tools have limited clinical utility in the assessment of self-harm.

Strengths and limitations

This is one of the few studies to compare widely used risk scales following self-harm in a 'head-to-head' prospective cohort study. The risk scales were administered by treating clinicians and prospectively evaluated in a large real-world sample of patients referred to liaison psychiatric services for self-harm. We used clear consistent terminology across sites and had near-complete patient follow-up, although it is possible that some patients could have moved or died during the study period without the knowledge

Table 2 Diagnostic accuracy statistics with 95% confidence intervals for *a priori* cut-off points

Scales, cut-off	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Positive predictive value, % (95% CI)	Negative predictive value, % (95% CI)	Likelihood ratio, positive (95% CI)	Likelihood ratio, negative (95% CI)	Diagnostic OR (95% CI)
Manchester Self-Harm Rule, 0/1+	97 (93–99)	20 (16–24)	34 (29–38)	94 (86–98)	1.2 (1.2–1.2)	0.1 (0.1–0.3)	8.7 (3.1–24.4)
ReACT Self-Harm Rule, 0/1+	97 (92–99)	23 (19–28)	35 (31–39)	95 (87–97)	1.3 (1.3–1.3)	0.2 (0.1–0.2)	8.5 (3.4–21.6)
SAD PERSONS scale							
0–4/5–6	16 (11–23)	90 (86–93)	40 (28–53)	72 (67–76)	1.6 (1.0–2.6)	0.9 (0.9–0.9)	1.7 (0.9–2.9)
5–6/7–10	1 (0–5)	99 (96–99)	25 (07–59)	70 (66–74)	0.8 (0.0–1.0)	1.0 (0.1–1.0)	0.8 (0.2–4.0)
Modified SAD PERSONS scale							
0–5/6–8	28 (21–36)	79 (74–83)	36 (27–45)	72 (67–77)	1.3 (1.1–1.6)	0.9 (0.8–1.0)	1.5 (0.9–2.3)
6–8/8+	1 (1–7)	98 (96–99)	13 (2–47)	70 (66–74)	0.3 (0.0–3.0)	1.0 (1.0–1.0)	0.3 (0.0–2.7)
Clinician global scale,[a] 0–5/6+	74 (66–80)	64 (59–69)	47 (41–53)	85 (80–90)	2.1 (2.0–2.1)	0.4 (0.4–0.4)	5.0 (3.2–7.8)
Patient global scale,[a] 0–5/6+	69 (61–77)	63 (57–68)	44 (37–50)	83 (78–87)	1.9 (1.8–1.9)	0.5 (0.5–0.5)	3.9 (3.3–5.9)
Barratt Impulsiveness Scale,[b] 0–96/97+	3 (1–8)	98 (96–99)	42 (19–68)	70 (67–74)	1.7 (0.0–1.8)	1.0 (0.9–1.0)	1.7 (0.5–5.4)

a. Mid-point cut off.
b. Cut-off used by Randall *et al.*[20]

of the clinical services. We used a broad definition of suicidal behaviour consistent with UK research and clinical practice. In fact, a *post hoc* analysis involving the 357 self-poisoning episodes (which would be consistently included in most definitions of suicidal behaviour internationally) generated similar results.

There is a risk of sampling bias as patients who refused to complete the research assessments or who were deemed inappropriate to participate were not included in the study, which may affect the generalisability of the results. Recent large multi-centre studies of self-harm in England suggest our sample was similar to overall patient samples in terms of gender,[19,41,50,51] method of self-harm,[3,19,41,50] and age.[41,50,51] Our recruitment rates are comparable with trials that involve obtaining individual consent from patients who have self-harmed.[6,28] However, the proportion of the sample with a prior history of self-harm (74.3%) and the repetition rate in our sample within 6 months (30%) was high,[40] possibly suggesting comparatively high levels of underlying morbidity and need.

The results of our study should be applicable to other psychiatric services with a similar case mix and incidence of repeat self-harm but may not be applicable to people who present to emergency departments and do not receive a psychiatric assessment. Patients who do not receive a psychiatric assessment following self-harm are likely to be younger, unemployed and use self-cutting as a method of self-harm, and have an absence of factors that indicate current mental illness.[32] Our results may also not be generalisable to people who engage in self-harm but do not present to hospital.

The repetition rate in these groups is likely to be lower, so it may be that the predictive performance of scales will be even worse. Our findings will also not be generalisable to patients who do not wait for assessment, who may be at higher risk than other patients who have self-harmed. Although we had a large sample from geographically dispersed sites this may not be representative of other hospitals in England or internationally. The five centres included in this study have an interest in self-harm management and research, and clinicians working in these sites may not be typical of those practising elsewhere. This could affect the generalisability of the results to other services, perhaps particularly those findings based on clinicians' global risk assessments.
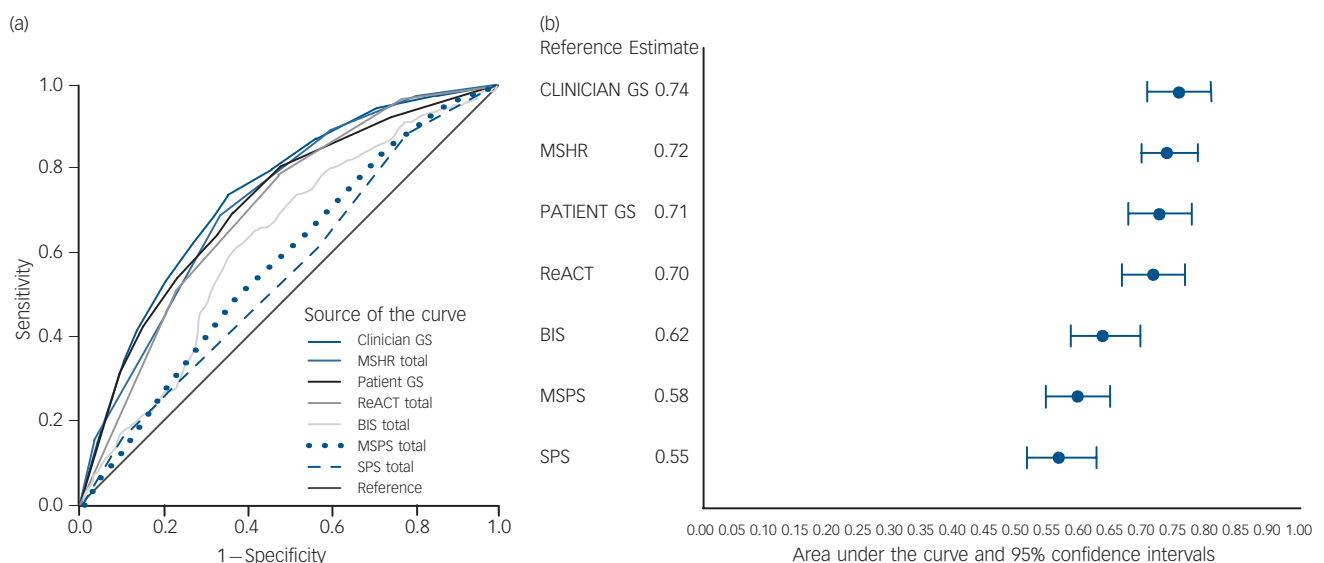

Fig. 1 The receiver operator characteristic curves (a) show the relationship between the proportion of true positives (sensitivity) and the proportion of false positives for the seven scales. The forest plot (b) shows the area under the curve estimates and 95% confidence intervals for the scales.

Clinician GS, clinician global scale; MSHR, Manchester Self-Harm Rule; Patient GS, patient global scale; ReACT, ReACT Self-Harm Rule; BIS, Barratt Impulsiveness Scale; MSPS, Modified SAD PERSONS Scale; SPS, SAD PERSONS Scale.

Table 3 Diagnostic accuracy statistics with 95% confidence intervals at optimal cut-off points using Youden's *J* Index

Scales, cut-off	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Positive predictive value, % (95% CI)	Negative predictive value, % (95% CI)	Likelihood ratio, + (95% CI)	Likelihood ratio, − (95% CI)	Diagnostic OR (95% CI)
Manchester Self-Harm Rule, 0–3/4+	69 (61–76)	66 (61–71)	47 (40–53)	84 (78–87)	2.0 (2.0–2.1)	0.5 (0.5–0.5)	4.4 (2.9–6.6)
ReACT Self-Harm Rule, 0–2/3+	79 (71–86)	52 (47–57)	41 (36–47)	85 (80–89)	1.6 (1.6–1.7)	0.4 (0.4–0.4)	4.0 (2.5–6.3)
SAD PERSONS scale, 0–2/3+	88 (82–93)	22 (18–27)	33 (28–37)	81 (72–88)	1.1 (1.1–1.1)	0.5 (0.4–0.7)	2.1 (1.2–3.7)
Modified SAD PERSONS scale, 0–5/6+	50 (42–57)	62 (57–67)	36 (30–43)	74 (69–79)	1.3 (1.3–1.4)	0.8 (0.8–0.8)	1.6 (1.1–2.4)
Clinician global scale, 0–5/6+	74 (66–80)	64 (59–69)	47 (41–53)	85 (80–89)	2.1 (2.0–2.1)	0.4 (0.4–0.4)	5.0 (3.3–7.7)
Patient global scale, 0–5/6+	70 (62–77)	63 (58–68)	45 (38–51)	83 (78–87)	1.9 (1.8–1.9)	0.5 (0.5–0.5)	3.9 (2.6–5.9)
Barratt Impulsiveness Scale, 0–75/76+	63 (55–70)	60 (55–66)	41 (34–46)	79 (74–84)	1.6 (1.5–1.6)	0.6 (0.6–0.6)	2.6 (1.7–3.9)

This observational study was designed to mimic how risk scales would be completed in clinical practice. The effect of ordering is likely to be minimal as the items from the scales (Manchester Self-Harm Rule, ReACT Self-Harm Rule, SAD PERSONS scale, Modified SAD PERSONS scale) were extracted from the notes subsequent to the information already gathered from the assessment. Although clinicians were masked to the scoring of the scales and the scale results, their use could have changed patient management. For example, patients deemed at greater risk might have been offered more intensive interventions that may have meant that they were then less likely to repeat self-harm. This could lead to an underestimate of predictive performance – that is, our findings on the risk scales might be unnecessarily pessimistic. We do not think this will have had a major impact on our findings. The availability of suitable interventions following self-harm is poor [6,52] and even if patients receive them, the evidence from randomised trials is that effect sizes are relatively modest.[6,53]

We focused on episodes rather than individuals in order to reflect the clinical reality of fluctuating risk[19] and to be consistent with national guidance that suggests each episode should be treated in its own right.[6] A small number of individuals contributed more than one assessment, which could potentially inflate the repetition rate and diagnostic accuracy statistics. There was a small decrease in the repetition rate for individuals when compared with the episode-based repetition rate (28.2% *v.* 30%), but this had little impact on our results. The order of the scales in terms of AUC was unchanged. The AUC values themselves were slightly attenuated (clinician global scale: AUC = 0.73; Manchester Self-Harm Rule: AUC = 0.71; patient global scale: AUC = 0.70; ReACT Self-Harm Rule AUC = 0.69, Barratt Impulsiveness Scale AUC = 0.62; Modified SAD PERSONS scale AUC = 0.59; and the SAD PERSONS scale: AUC = 0.56).

Our outcome was repeated self-harm rather than suicide. Suicide is a critically important outcome for research and clinical practice but a challenging area for diagnostic accuracy studies. The low base rate makes prediction difficult and the diagnostic accuracy of scales is generally poor.[6] Future research could examine the performance of risk scales in the prediction of suicide, which may be possible with very large multicentre prospective cohort studies. However, factors associated with future suicide may not be the same as those that are associated with risk of future self-harm.[6]

Comparison with previous research

The performance of the scales using *a priori* cut-off points is consistent with previous research. The Manchester Self-Harm Rule had similar sensitivity (97%) and specificity (20%) to previous validation studies of the Manchester Self-Harm Rule,[19] but lower specificity than in the original study.[17] The ReACT Self-Harm Rule had an equivalent sensitivity (97%) but higher specificity (23%) in this sample than in the original study (97% and 20%, respectively).[19]

Using the same cut-off as Randall *et al*,[20] the Barratt Impulsiveness Scale had a lower sensitivity (3% *v.* 20%) in this sample but similar specificity (98% *v.* 97%, respectively). The poor performance of the Barratt Impulsiveness Scale may be a result of cultural differences as the scale was developed in the USA and may not be directly relevant for a clinical population in the UK; for example, the item 'I squirm at plays and lectures' caused queries and some respondents found it difficult to answer.[13] The length of the scale (30 items) may also have been an issue. It should also be noted that this scale was not developed as an instrument to predict repeat suicidal behaviour but as a measure of impulsivity, which is just one risk factor for suicidal behaviour.[54] The poor performance of the SAD PERSONS scale and Modified SAD PERSONS scale in predicting repeat self-harm is consistent with previous cohort studies.[15,55]

Clinical implications

The use of risk scales is dependent on clinical context.[13] For example, clinicians may prefer scales with high sensitivity for screening or ruling out a risk of a condition, or scales high in specificity for later stages of assessment or ruling in patients for treatment.[13] However, our findings suggest that risk scales on their own have little role in the management of suicidal behaviour. For example, one of the best performing scales, the Manchester Self-Harm Rule, captured 97 out of every 100 repeat episodes, but incorrectly classified 80/100 of episodes that did not lead to repetition as high risk. Of 100 episodes rated as high risk only 30 resulted in repetition. The scales performed no better (and in some cases significantly worse) than simply asking clinicians or patients what they thought of the future risk. The usefulness of the scales might improve if the cut-off points were tailored to local clinical settings, but the results would then not be generalisable and the cut-offs may not be stable over time.

It was perhaps surprising that the crude global estimates of risk performed comparatively well. On the other hand, clinicians in this study were generally experienced and may have used all the available clinical information and direct observation to come to a more balanced judgement of risk than a score on a simple scale. Of course, it could also be that the clinicians used the scales themselves to inform their overall judgement, but they were not provided with the scoring schedule for the scales and much of the data consisted of items that would be collected as part of routine assessment. These explanations would of course not apply to the patient assessment of risk.

Is there scope for using global estimation of risk scales as a useful part of routine assessments? We think this is unlikely since the positive predictive values for the clinician global scale

indicated that for every 100 patients rated as high risk, fewer than half would go on to repeat. Of 100 patients who did not repeat, 36 would be incorrectly classified as at high risk.

Of course, risk scales might be useful in ways other than prediction. For example, to help structure assessments, to ensure crucial items are not missed or as measures of change. Can risk scales do any harm? Some observational evidence suggests that routine aspects of clinical care such as psychosocial assessment and psychiatric admission could contribute to a reduction in risk.[28,30,56] Risk scales may have a negative impact on the beneficial aspects of routine psychosocial assessments.[11,57] They may be perceived as a negative tokenistic 'tick box' exercise by both clinicians and patients and erode the potential to collaboratively evaluate risk of future self-harm and determine appropriate management.[58] At a time of increased service pressures it might even be argued that the use of risk scales to determine patient management actually wastes valuable resources.[8]

Future research

Consistent with clinical guidelines, our data suggest that risk scales should not be used to determine patient management or risk of future self-harm. [6] One relatively unexplored area is the use of risk assessment as an intervention. In forensic settings, randomised trials of formal risk assessment have had conflicting results.[59,60] Randomised controlled trials could test the impact of using risk scales v. assessment as usual on patient management and repeat self-harm, including adverse events and cost-effectiveness. We are currently undertaking health economic modelling work that will provide an indication of how good risk tools might need to be in terms of predictive ability in order to be cost-effective.

Given the poor performance of scales, it is possible that the scales may be missing important aspects relevant to repeat suicidal behaviour (for example social, cultural, economic or psychological processes).[56,61,62] Future research should include patients in the development of appropriate measures and assessments and could also consider suicide as an outcome. It is likely that the predictive ability of assessment varies according to clinician factors (such as level of experience, professional background), patient factors (history of suicidal behaviour or psychiatric treatment), and assessment factors (received a psychiatric assessment or did not wait for assessment), and these could also be investigated. Studies might examine the role of global clinician and patient assessments of risk, with a focus on their predictive performance but also an examination of the factors that contribute to these complex judgements.

Leah Quinlivan, PhD, **Jayne Cooper**, PhD, Centre for Suicide Prevention, Manchester Academic Health Science Centre, University of Manchester, Manchester; **Declan Meehan**, RMN, **Damien Longson**, FRCPsych, Greater Manchester Mental Health and NHS Foundation Trust, Manchester; **John Potokar**, MRCPsych, Avon & Wiltshire Mental Health Foundation Trust, Bristol, University Hospitals Bristol, NHS Foundation Trust, Bristol and School of Social and Community Medicine, University of Bristol, Bristol; **Tom Hulme**, MSc, University Hospitals Bristol, NHS Foundation Trust, Bristol, Bristol; **Jennifer Marsden**, BA, Derbyshire Healthcare NHS Foundation Trust, Derby; **Fiona Brand**, RMN, **Kezia Lange**, MRCPsych, Oxford Health NHS Foundation Trust, Oxford; **Elena Riseborough**, RMN, **Lisa Page**, PhD, Sussex Partnership NHS Foundation Trust, Worthing; **Chris Metcalfe**, PhD, School of Social and Community Medicine, University of Bristol, Bristol; **Linda Davies**, Institute of Population Health, University of Manchester, Manchester; **Rory O' Connor**, PhD, Institute of Health and Wellbeing, University of Glasgow, Glasgow; **Keith Hawton**, DSc, Centre for Suicide Research, University Department of Psychiatry, Warneford Hospital, Oxford; **David Gunnell**, DSc, School of Social and Community Medicine, University of Bristol, Bristol; **Nav Kapur**, FRCPsych, Centre for Suicide Prevention, Manchester Academic Health Science Centre, University of Manchester, and Greater Manchester Mental Health NHS Foundation Trust, Manchester, UK

Correspondence: Leah Quinlivan, Jean McFarlane Building, University of Manchester, Oxford Road, Manchester M13 9PL, UK. Email: leah.quinlivan@manchester.ac.uk

First received 30 Jun 2016, final revision 5 Oct 2016, accepted 13 Nov 2016

References

1 Bergen H, Hawton K, Waters K, Cooper J, Kapur N. Psychosocial assessment and repetition of self-harm: the significance of single and multiple repeat episode analyses. *J Affect Disord* 2010; **127**: 257–65.

2 Bergen H, Hawton K, Waters K, Ness J, Cooper J, Steeg S, et al. Premature death after self-harm: a multicentre cohort study. *Lancet* 2012; **380**: 1568–74.

3 Hawton K, Bergen H, Cooper J, Turnbull P, Waters K, Ness 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.

4 Haw C, Bergen H, Casey D, Hawton K. Repetition of deliberate self-harm: a study of the characteristics and subsequent deaths in patients presenting to a general hospital according to extent of repetition. *Suicide Life Threat Behav* 2007; **37**: 379–96.

5 Zahl DL, Hawton K. Repetition of deliberate self-harm and subsequent suicide risk: long-term follow-up study of 11583 patients. *Br J Psychiatry* 2004; **185**: 70–5.

6 National Institute for Health and Care Excellence. *Self-harm. The NICE Guideline on Longer-Term Management. National Clinical Guideline Number 133*. The British Psychological Society and The Royal College of Psychiatrists, 2011.

7 Royal Australian and New Zealand College of Psychiatrists Clinical Practice Guidelines Team for Deliberate Self-harm. Australian and New Zealand clinical practice guidelines for the management of adult deliberate self-harm. *Aust NZ J Psychiatry* 2004; **38**: 868–84.

8 Bolton JM, Gunnell D, Turecki G. Suicide risk assessment and intervention in people with mental illness. *BMJ* 2015; **35**: h4978.

9 Kapur N, Murphy E, Cooper J, Bergen H, Hawton K, Simkin S, et al. Psychosocial assessment following self-harm: results from the multi-centre monitoring of self-harm project. *J Affect Disord* 2008; **106**: 285–93.

10 Hawton K. Psychiatric assessment and management of deliberate self-poisoning patients. *Medicine* 2016; **44**: 103–5.

11 Royal College of Psychiatrists. *Self-Harm, Suicide, and Risk: Helping People who Self-Harm. Final Report of a Working Group*. Royal College of Psychiatrists, 2010.

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

13 Quinlivan L, Cooper J, Davies L, Hawton K, Gunnell D, Kapur N, 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.

14 Bilén K, Ponzer S, Ottosson C, Castrén M, Owe-Larsson B, Ekdahl K, et al. Can repetition of deliberate self-harm be predicted? A prospective multicenter study validating clinical decision rules. *J Affect Disord* 2013; **149**: 253–8.

15 Bolton JM, Spiwak R, Sareen J. Predicting suicide attempts with the SAD PERSONS scale: a longitudinal analysis. *J Clin Psychiatry* 2012; **76**: e735–41.

16 Carter GL, Clover KA, Bryant JL, Whyte IM. Can the Edinburgh Risk of Repetition Scale predict repetition of deliberate self-poisoning in an Australian clinical setting? *Suicide Life Threat Behav* 2002; **32**: 230–9.

17 Cooper J, Kapur N, Dunning J, Guthrie E, Appleby L, Mackway-Jones K. A clinical tool for assessing risk after self-harm. *Ann Emerg Med* 2006; **48**: 459–66.

18 Spittal MJ, Pirkis J, Miller M, Carter G, Studdert DM. 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.

19 Steeg S, Kapur N, Webb R, Applegate E, Stewart SL, Hawton K, 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.

20 Randall JR, Rowe BH, Colman I. Emergency department assessment of self-harm risk using psychometric questionnaires. *Can J Psychiatry* 2012; **57**: 21.

21 Waern M, Sjöström N, Marlow T, Hetta J. Does the Suicide Assessment Scale predict risk of repetition? A prospective study of suicide attempters at a hospital emergency department. *Eur Psychiatry* 2010; **25**: 421–6.

22 Hockberger RS, Rothstein RJ. Assessment of suicide potential by nonpsychiatrists using the SAD PERSONS score. *J Emerg Med* 1988; **6**: 99–107.

23 Patton JH, Stanford MS. Factor structure of the Barratt Impulsiveness Scale. *J Clin Psychol* 1995; **51**: 768–77.

24 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.

25 Larkin C, Di Blasi Z, Arensman E. Risk factors for repetition of self-harm: a systematic review of prospective hospital-based studies. *PLoS ONE* 2014; **9**: e84282.

26 Chan MK, Bhatti H, Meader N, Stockton S, Evans J, O'Connor RC, et al. Predicting suicide following self-harm: systematic review of risk factors and risk scales. *Br J Psychiatry* 2016; **209**: 277–83.

27 Bewick V, Cheek L, Ball J. Statistics review 13: receiver operating characteristic curves. *Critical Care* 2004; **8**: 508.

28 Kapur N, Steeg S, Webb R, Haigh M, Bergen H, Hawton K, et al. Does clinical management improve outcomes following self-harm? Results from the multicentre study of self-harm in England. *PLoS ONE* 2013; **8**: e70434.

29 Majid M, Tadros M, Tadros G, Singh S, Broome MR, Upthegrove R, et al. Young people who self-harm: a prospective 1-year follow-up study. *Soc Psychiatry Psychiatr Epidemiol* 2016; **51**: 171–81.

30 Kapur N, Steeg S, Turnbull P, Webb R, Bergen R, Hawton K, et al. Hospital management of suicidal behaviour and subsequent mortality: a prospective cohort study. *Lancet Psychiatry* 2015; **2**: 809–16.

31 Steeg S, Haigh M, Webb RT, Kapur N, Awenat Y, Gooding P, et al. The exacerbating influence of hopelessness on other known risk factors for repeat self-harm and suicide. *J Affect Disord* 2016; **190**: 522–8.

32 Kapur N, Murphy E, Cooper J, Bergen H, Hawton K, Simkin S, et al. Psychosocial assessment following self-harm: results from the multi-centre monitoring of self-harm project. *J Affect Disord* 2008; **106**: 285–93.

33 Silverman MM, Berman AL, Sanddal ND, O'Carroll PW, Joiner TE. Rebuilding the Tower of Babel: a revised nomenclature for the study of suicide and suicidal behaviors Part 1: background, rationale, and methodology. *Suicide Life Threat Behav* 2007; **37**: 248–63.

34 O'Carroll PW, Berman AL, Maris RW, Moscicki EK, Tanney BL, Silverman MM. Beyond the Tower of Babel: a nomenclature for suicidology. *Suicide Life Threat Behav* 1996; **26**: 237–52.

35 Andover MS, Morris BW, Wren A, Bruzzese ME. The co-occurrence of non-suicidal self-injury and attempted suicide among adolescents: distinguishing risk factors and psychosocial correlates. *Child Adolesc Psychiatry Ment Health* 2012; **6**: 11.

36 Owens D, Kelley R, Munyombwe T, Bergen H, Hawton K, Cooper J, et al. Switching methods of self-harm at repeat episodes: findings from a multicentre cohort study. *J Affect Disord* 2015; **180**: 44–51.

37 Clements C, Jones S, Morriss R, Peters S, Cooper J, While D, et al. Self-harm in bipolar disorder: findings from a prospective clinical database. *J Affect Disord* 2015; **173**: 113–9.

38 Hawton K, Zahl D, Weatherall R. Suicide following deliberate self-harm: long-term follow-up of patients who presented to a general hospital. *Br J Psychiatry* 2003; **182**: 537–42.

39 Owens D, Horrocks J, House A. Fatal and non-fatal repetition of self-harm. *Br J Psychiatry* 2002; **181**: 193–9.

40 Carroll R, Metcalfe C, Gunnell D. Hospital presenting self-harm and risk of fatal and non-fatal repetition: systematic review and meta-analysis. *PLoS ONE* 2014; **9**: e89944.

41 Cooper J, Kapur N, Mackway-Jones K. A comparison between clinicians' assessment and the Manchester Self-Harm Rule: a cohort study. *Emerg Med J* 2007; **24**: 720–1.

42 Patterson WM, Dohn HH, Bird J, Patterson GA. Evaluation of suicidal patients: the SAD PERSONS scale. *Psychosomatics* 1983; **24**: 343–9.

43 Kapur N, Cooper J, Hiroeh U, May C, Appleby L, House A. Emergency department management and outcome for self-poisoning: a cohort study. *Gen Hosp Psychiatry* 2004; **26**: 36–41.

44 Cooper J, Steeg S, Bennewith O, Lowe M, Gunnell D, House A, et al. Are hospital services for self-harm getting better? An observational study examining management, service provision and temporal trends in England. *BMJ Open* 2013; **3**: e003444.

45 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**: 177–88.

46 Higgins JP, Green S. *Cochrane Handbook for Systematic Reviews Of Interventions (vol 5)*. Wiley Online Library, 2008.

47 Hosmer Jr DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression (vol 398)*. John Wiley & Sons, 2013.

48 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44**: 837–45.

49 Allison PD. *Missing Data (vol 136)*. Sage Publications, 2001.

50 Ness J, Hawton K, Bergen H, Cooper J, Steeg S, Kapur N, et al. Alcohol use and misuse, self-harm and subsequent mortality: an epidemiological and longitudinal study from the multicentre study of self-harm in England. *Emerg Med J* 2015; 32: 793–9.

51 Geulayov G, Kapur N, Turnbull P, Clements C, Waters K, Ness J, et al. Epidemiology and trends in non-fatal self-harm in three centres in England, 2000-2012: findings from the Multicentre Study of Self-Harm in England. *BMJ Open* 2016; **6**: e010538.

52 Hawton K, Witt KG, Taylor Salisbury TL, Arensman E, Gunnell D, Hazell P, et al. *Psychosocial Interventions for Self-Harm in Adults*. The Cochrane Library, 2016.

53 O'Connor E, Gaynes BN, Burda BU, Soh C, Whitlock EP. Screening for and treatment of suicide risk relevant to primary care: a systematic review for the US Preventive Services Task Force. *Ann Antern Med* 2013; **158**: 741–54.

54 O'Connor RC, Nock MK. The psychology of suicidal behaviour. *Lancet Psychiatry* 2014; **1**: 73–85.

55 Saunders K, Brand F, Lascelles K, Hawton K. The sad truth about the SADPERSONS Scale: an evaluation of its clinical utility in self-harm patients. *Emerg Med J* 2014; **31**: 796–8.

56 Carroll R, Metcalfe C, Steeg S, Davies NM, Cooper J, Kapur N, et al. Psychosocial assessment of self-harm patients and risk of repeat presentation: an instrumental variable analysis using time of hospital presentation. *PLoS ONE* 2016; **11**: e0149713.

57 Smith MJ, Bouch J, Bradstreet S, Lakey T, Nightingale A, O'Connor RC. Health services, suicide, and self-harm: patient distress and system anxiety. *Lancet Psychiatry* 2015; **2**: 275–80.

58 Hunter C, Chantler K, Kapur N, Cooper J. Service user perspectives on psychosocial assessment following self-harm and its impact on further help-seeking: a qualitative study. *J Affect Disord* 2013; **145**: 315–23.

59 Abderhalden C, Needham I, Dassen T, Halfens R, Haug HJ, Fischer JE. Structured risk assessment and violence in acute psychiatric wards: randomised controlled trial. *Br J Psychiatry* 2008; **193**: 44–50.

60 Troquete NA, van den Brink RH, Beintema H, Mulder T, van Os TW, Schoevers RA, et al. Risk assessment and shared care planning in out-patient forensic psychiatry: cluster randomised controlled trial. *Br J Psychiatry* 2013; **202**: 365–71.

61 Coope C, Donovan J, Wilson C, Barnes M, Metcalfe C, Hollingworth W, et al. Characteristics of people dying by suicide after job loss, financial difficulties and other economic stressors during a period of recession (2010–2011): a review of coroners records. *J Affect Disord* 2015; **183**: 98–105.

62 Haw C, Hawton K, Gunnell D, Platt S. Economic recession and suicidal behaviour: Possible mechanisms and ameliorating factors. *Int J Soc Psychiatry* 2015; **61**: 73–81.

Online supplement DS1

Risk scales tested in the predicting risk of repeat self-harm cohort study

Scales	Scale details		Assessment method	Items and cut-off points
Manchester Self-Harm Rule[17]	4 items • History of self-harm • Prior psychiatric treatment • Benzodiazepines • Current psychiatric treatment		Taken from clinical interview and notes post assessment	Score of any one indicates moderate/high risk
The ReACT Self-Harm Rule[19]	4 items • Recent self-harm (last year) • Cutting as method • Lives alone or homeless • Current psychiatric treatment		Taken from clinical interview and notes post assessment	Score of any one indicates moderate/high risk
The SAD PERSONS scale[15,42]	10 items • Male gender • Older age • Depression • Previous suicide attempt	• Excess alcohol or substance use • Rational thinking loss • Social supports lacking • Organised plan • No spouse • Sickness	Taken from clinical interview and assessment notes	3 categories of risk: 0–4, 5–6, 7–10 for low, moderate, and high respectively
The Modified SAD PERSONS scale[15,22]	10 items • Male gender • >19<45 • Depression or hopelessness • Previous suicidal attempts or psychiatric care	• Rational thinking loss • Single, widowed or divorced • Organised or serious attempt • No social support	Taken from clinical interview and assessment notes	3 categories of risk, 0–5, 6–8, 9–14, for low, moderate, and high respectively

	• Excessive ethanol or drug use	• Stated future intent	
	•		
The Barratt Impulsiveness Scale[20,23]	30-items based on personality. Responses scored on 4-point likert scale.	Patient completed	30 item self-report. Higher scores indicate greater impulsivity. A cut-off of 97 was used to denote high risk based on Randall et al (2012)[20]
Patient global estimation of risk scale	1–10 likert scale evaluating likelihood of risk of repeat self-harm in the next six months	Patient completed	1–10 scale, mid-point (0–5, 6+) used as the cut-off
Clinician global estimation of risk scale	1- 10 likert scale evaluating likelihood of risk of repeat self-harm in the next six months	Clinicians completed as part of the clinical assessment	1–10 scale, mid-point (0–5, 6+) used as the cut-off

Online supplement DS2

Definitions[13,27,63]

Sensitivity (Sens) – the proportion of people who repeat self-harm and are correctly identified by the scale as high risk

Specificity (Spec) – the proportion of people who do not repeat self-harm and are correctly identified by the scale as low risk

Positive predictive value (PPV) –The probability that the person identified as high risk by the scale actually went on to repeat self-harm

Negative predictive value (NPV) –The probability that the person identified as low risk by the scale actually did not repeat self-harm

Positive Likelihood ratio (LR+) –The increased likelihood of a high-risk scale result in a patient who repeats self-harm versus one who does not

Negative likelihood ratio (LR-) –The decreased likelihood of a low-risk scale result in a patient who repeats self-harm versus one who does not

Diagnostic odds ratio (DOR) –The odds of a high-risk result in a patient who repeats self-harm versus one who does not (interpreted the same as an odds ratio).

Receiver operating characteristic curve (ROC): Graphically shows the overall discrimination ability of a scale to identify patients who repeat self-harm compared with those who do not at various cut-off points (plotted as sensitivity versus 1-specifcity). The performance of the scale is indicated by the calculation of the area under the curve (AUC). Higher AUC indicate greater discriminatory power.

Youden´s J index (J): The difference between true positive rate and false positive rate. It provides the maximum point on the curve for both sensitivity and specificity

Additional reference

63 Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology* 2005:7 3–81.