Review article

Predicting suicidal behaviours using clinical instruments: systematic review and meta-analysis of positive predictive values for risk scales[†]

Gregory Carter, Allison Milner, Katie McGill, Jane Pirkis, Nav Kapur and Matthew J. Spittal

Background
Prediction of suicidal behaviour is an aspirational goal for clinicians and policy makers; with patients classified as 'high risk' to be preferentially allocated treatment. Clinical usefulness requires an adequate positive predictive value (PPV).

Aims
To identify studies of predictive instruments and to calculate PPV estimates for suicidal behaviours.

Method
A systematic review identified studies of predictive instruments. A series of meta-analyses produced pooled estimates of PPV for suicidal behaviours.

Results
For all scales combined, the pooled PPVs were: suicide 5.5% (95% CI 3.9–7.9%), self-harm 26.3% (95% CI 21.8–31.3%) and self-harm plus suicide 35.9% (95% CI 25.8–47.4%). Subanalyses on self-harm found pooled PPVs of 16.1% (95% CI 11.3–22.3%) for high-quality studies, 32.5% (95% CI

26.1–39.6%) for hospital-treated self-harm and 26.8% (95% CI 19.5–35.6%) for psychiatric in-patients.

Conclusions
No 'high-risk' classification was clinically useful. Prevalence imposes a ceiling on PPV. Treatment should reduce exposure to modifiable risk factors and offer effective interventions for selected subpopulations and unselected clinical populations.

Declaration of interest
N.K. chaired the NICE guidelines for the longer term management of self-harm in England but the views in this paper are the author's own and not those of NICE or the Department of Health (UK). G.C. chaired the Royal Australian and New Zealand College of Psychiatrists' (RANZCP's) Clinical Practice Guidelines for Deliberate Self Harm but the views in this paper are the author's own and not those of the RANZCP.

Copyright and usage
© The Royal College of Psychiatrists 2017.

Mental health clinicians treat patients who are at much greater risk of suicide, suicide attempts or non-fatal self-harm than the general population.[1,2] Clinicians would like to be able to predict with acceptable accuracy, for a clinically meaningful time frame, which individual patients will subsequently die by suicide or have a further episode of non-fatal self-harm so that preventive interventions can be preferentially allocated to those classified as 'high risk' for those outcomes.[3] Historically, there have been three generations of prediction approaches: unassisted clinician prediction (first), standardised scales or biological tests (second) and scales derived from statistical modelling (third). Many clinical instruments have been utilised for prediction including: psychological scales such as versions of the Beck Depression Inventory (BDI)[4] or the SADPERSONS scale;[5] biological tests such as the dexamethasone suppression test (DST)[6] and the cerebrospinal fluid (CSF) 5-hydroxyindoleacetic acid (5-HIAA) concentration test;[7] and scales derived from statistical models such as the ReACT Self-Harm Rule[8] and the Repeated Episodes of Self-Harm (RESH) score.[9]

At the policy level, the use of risk assessment classification to determine treatment allocation has been strongly endorsed in the USA. The (US) National Action Alliance for Suicide Prevention's Research Prioritization Task Force has made a recommendation to 'find ways to assess who is at risk for attempting suicide in the immediate future'. This recommendation is specifically 'related to the task of identifying and predicting near-term suicide risk at the individual patient level'.[10] Similarly, the (US) National Strategy for Suicide Prevention stated the need to 'Fund the development of suicide screening and assessment tools that will be non-proprietary and widely available' (Objective 7.4); and 'Develop standardized protocols for use within emergency departments based on common clinical presentation to allow for more differentiated responses

based on risk profiles and assessed clinical needs' (Objective 9.6).[11] There have been clear objections about the clinical utility of this approach, based on the inaccuracy of predictive 'tests' used as the basis for allocation of treatment.[12] In the UK, the National Institute for Health and Care Excellence (NICE) guidelines have instead suggested 'Do not use risk assessment tools and scales to predict future suicide or repetition of self-harm' and emphasised a shift in recommendations from 'risk assessment' to 'needs assessment' to determine allocation of clinical aftercare.[13] The relevant accuracy statistics for clinicians are the positive predictive value (PPV) and the negative predictive value (NPV) of a test; and as a basis for allocation of treatment, the PPV is the key statistic. Simply put: 'The positive predictive value . . . expresses the proportion of those with positive test results who truly have disease'.[14] Unlike sensitivity and specificity, the PPV and NPV are highly dependent on the prevalence of the outcome of interest, which means that the values for these measures are not simply transferable from one clinical population to another with different prevalence of disease.[14] A few systematic reviews of predictive instruments have reported sensitivity and specificity ranges for specific tests or scales.[15,16] A recent review explored a wide range of diagnostic accuracy measures for a small number of risk scales used to assess patients after presentation for self-harm.[17] However, there have been no meta-analyses to produce pooled estimates for the PPV for predictive instruments in mental health patient populations.

Method

Key questions

Our key question for the review was: is the classification of mental health patients as being 'high risk' for subsequent suicide death or

self-harm (for example non-fatal self-harm, deliberate self-harm, self-harm, suicide attempt or parasuicide), by risk assessment, using either psychological scales, biological tests or third-generation scales, sufficiently accurate for clinical use? Our subquestion was: what are the pooled estimates for PPV of those clinical risk assessments in clinical populations?

Databases and search terms used

The systematic review was conducted using the PRISMA statement and associated set of instructions.[18] The search terms used were selected from past reviews and included: synonyms for suicidal behaviours including suicide and non-fatal self-harm (for example "self$harm", "attempted suicide", "parasuicide", "self$injur*", "self$poison*", "suicide*"), synonyms for repetition (for example "repeat*", "recur*", "re$present*", "recidiv*") and synonyms for cohort study (for example "follow$up", "retrospective", "predict*", "prospect*", "longitudinal"). The databases used for the search included Medline, PsychInfo, Embase, CINHAL, Web of Science, Cochrane trials and Scopus. No time limits were used. We also hand searched key journals in the field; reviewed the reference lists of each paper retrieved; and used the 'find similar' and 'find citing' functions for seminal papers in Web of Science and PubMed. We contacted corresponding authors to provide clarification of results when these were unclear.

Inclusion and exclusion criteria

Studies were eligible for inclusion if: (a) they used a longitudinal cohort design; (b) they reported on a psychological scale, a biological test or a third-generation scale; (c) the scale was used as a risk assessment tool by using a cut-off score to classify participants as being at 'high risk' for subsequent suicidal behaviour; and (d) they reported data for suicide or self-harm outcomes during a follow-up period. There was no restriction based on study population, setting or age group. Only studies published in English were included. There was no restriction based on the time period when the study was conducted.

Studies were excluded if they did not: (a) use a clinical predictive scale of some type (for example unassisted clinician opinion) to 'predict' suicidal behaviours; (b) provide the minimum necessary extractable data for the meta-analyses; (c) did not have information for suicidal behaviour outcomes during a specified follow-up period; or (d) reported data from subsamples reported in other studies.

Data collection process

Three of the authors (K.M., A.M., M.J.S.) extracted descriptive information for each study. Individual studies could report on more than one scale, so we extracted the name of each scale and the cut-point used to predict outcomes (suicide, self-harm or self-harm plus suicide). For the meta-analyses, for each scale we used the 2 × 2 contingency tables or, if not available, we used the reported sensitivity, specificity and prevalence to calculate the values of interest using Bayes' rule. These data were recorded on forms, which were piloted on five evaluations and then modified before final use by four authors working in pairs (G.L.C.–K.M., A.M.–M.J.S.). Data were extracted by two independent raters, non-agreement was settled by discussion and consensus and reviewed by a third rater if needed.

Ratings of bias

We used the QUADAS-2 tool (QUality Assessment of Diagnostic Accuracy Studies – version 2) to assess risk of bias in four

domains: patient selection (two items: participant selection (random or consecutive) and exclusions less than 15% of population), index test (two items: masking to outcome and pre-specified cut-points), reference standard (two items: classification of outcomes and masking of rating), and flow and timing (3 items: duration of follow-up 1 year or less, same outcome measurement for all, drop-out less than 15%).[19]

The QUADAS-2 forms were piloted on five evaluations and then modified before final use by two of the authors (G.L.C., A.M.) for each scale evaluation. Each item was phrased as a question requiring a rating of 'yes', 'no' or 'unclear' and each domain was then rated for risk of bias, classified as 'low', 'high' or 'unclear'. The ratings of risk of bias for the four domains were used to provide a pooled rating of risk of bias for all the scales included in the meta-analyses. A subgroup of scale evaluations were classified as high quality (i.e. low risk of bias) if the ratings in the two most important domains (patient selection and flow and timing) were rated as low risk; and this subgroup was used for meta-analysis.

Data analysis

We classified studies as reporting biological scales or psychological scales (including third-generation scales) or both. We reported study-specific descriptive results using the $n = 70$ studies as the unit of analysis and scale-specific descriptive and meta-analysis results using the $k = 128$ study-outcome-sample-scales as the unit of analysis (online Table DS1). This latter unit of analysis reflects the different levels of information within each study – information on the study itself, the outcomes explored within each study, any subsamples that were used (whole sample, training sample, validation sample), and finally information on each scale that was evaluated. We therefore refer to this unit of analysis as the scale. Because scale-specific PPV values are proportions, we used the binomial-normal model to estimate the pooled PPV.[20] This is a random-effects logistic regression model. The pooled PPV and the confidence intervals are estimated on the logit scale and then back transformed to a proportion for interpretation.

For the high-quality studies, third-generation studies, studies in hospital settings and studies of single psychological scales, there were a smaller number of studies, and so we combined self-harm and self-harm plus suicide outcomes (since the additional suicide events did not substantially inflate the prevalence of the self-harm outcome) as a composite outcome. The combined outcome of self-harm plus suicide can be interpreted as a self-harm outcome for the purposes of the estimated PPV.

We grouped scale evaluations by type and completed meta-analyses for: all scales combined (any suicidal behaviour, self-harm, self-harm plus suicide and suicide), high-quality evaluations (self-harm and self-harm plus suicide combined), all psychological scales (any suicidal behaviour, self-harm, self-harm plus suicide and suicide), all biological tests (any suicidal behaviour, self-harm, self-harm plus suicide and suicide) and third-generation scales (self-harm and self-harm plus suicide combined).

We also completed meta-analyses for individual scales where there were three or more evaluations available. These included biological tests, DST and CFS 5-HIAA levels (suicide); psychological scales, Buglass and Horton, SADPERSONS, Beck Hopelessness Scale (BHS), Beck Depression Inventory (BDI); and third-generation scales, Manchester Self-Harm Rule (MSHR), Edinburgh Risk Rating Scale (ERRS) (self-harm and self-harm plus suicide combined).

Publication bias was assessed using funnel plots for all studies and all scales combined, although it is acknowledged that this may be of limited usefulness in the meta-analyses of predictive studies.

All meta-analyses were performed using the metafor package[21] in R (version 3.20).[22]

Results

The search produced 32 166 articles (including duplicates). Keyword screening in title and abstract identified 1076. We removed 355 duplicates. We screened the abstracts of the remaining 721 articles, removed 510 and included a further 93 from other sources. We then assessed 304 articles by reading the full text (including one study that was reviewed twice because it reported on a psychological and a biological scale). A total of 233 articles were excluded from this set (135 did not predict suicidal behaviour or were not longitudinal and 98 had no extractable data); leaving 70 articles for analysis (Fig. 1).

Overview of studies and scales

From the 70 selected studies, 52 assessed psychological scales,[8,9,23–72] 17 biological measures[7,73–88] and one reported on both.[89] An overview of the studies is shown in online Tables DS1 and DS2. Studies came from North America (psychological $n = 24$, biological $n = 9$), UK (psychological $n = 13$), Europe (psychological $n = 12$, biological $n = 7$, both $n = 1$), Australia and New Zealand (psychological $n = 3$) and one where the country was not reported (biological $n = 1$). The earliest study was published in 1966[34] and the latest in 2014.[71] Publication of articles over time show a phasic distribution with peaks in the 1980s ($n = 18$), 2000s ($n = 19$) and a further peak since 2010 ($n = 17$ currently).

Settings and samples

Most studies recruited adults (psychological $n = 29$, biological $n = 13$, both $n = 1$), others combined youth and adults (psychological $n = 9$) or adolescents only (psychological $n = 4$) and some did not report ages (psychological $n = 10$, biological $n = 4$). The samples were typically drawn from patients with recent self-harm or suicide ideation (psychological $n = 33$, biological $n = 3$, both $n = 1$) or from psychiatric populations (psychological $n = 15$, biological $n = 14$) with a minority from other populations (psychological $n = 4$). For psychiatric populations, the specific disorders (where reported) were mood disorders (psychological $n = 1$, biological $n = 11$), first-episode psychosis or schizophrenia (psychological $n = 2$, biological $n = 1$), post-traumatic stress disorder (clinical $n = 1$) and personality disorder (biological $n = 1$). Other populations were military veterans (psychological $n = 1$, biological $n = 1$) and prisoners (psychological $n = 2$).

Follow-up time points

The follow-up periods varied from 6 months or less (psychological $n = 17$, biological $n = 1$) to more than 10 years (psychological $n = 3$, biological $n = 2$). The most common length of follow-up was 1 year (psychological $n = 13$, biological $n = 5$, both $n = 1$).

QUADAS quality ratings

In total, 17% of scales were judged as having low risk of bias for patient selection ($k = 22$), 49% for choice of index test ($k = 63$), 59% for the reference standard ($k = 76$) and 34% for flow and timing of patients ($k = 44$) (online Fig. DS1). In all, 16 scales were judged as high quality overall because of low risk of bias in patient selection and flow and timing. Details can be seen in online Fig. DS1.

Pooled estimates of PPV

The forest plots of the study-specific PPVs for all scales and for each suicidal behaviour are contained in the online Fig. DS2. For all scales and any suicidal behaviour combined ($k = 128$),
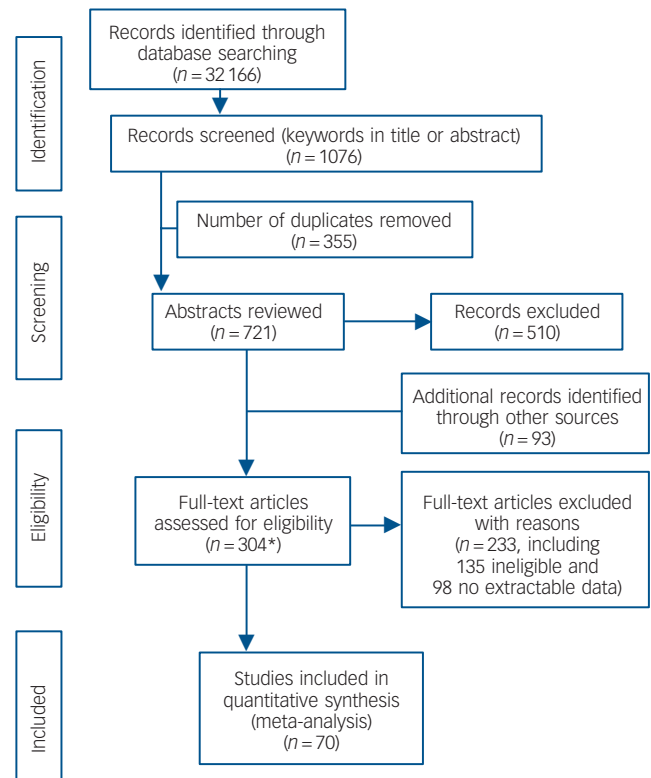

Fig. 1 PRISMA flow diagram.

*Includes one study that was assessed twice as it held data relevant to both a clinical and biological scale.

the overall pooled estimate PPV was 16.0%; for self-harm, ($k = 62$) 26.3%; for self-harm or self-harm and suicide combined ($k = 15$) 35.9%; and for suicide ($k = 51$) 5.5% (Fig. 2).

When restricted to high-quality evaluations ($k = 16$) for self-harm or self-harm plus suicide combined, the pooled PPV estimate was 16.1%. For the psychological instruments the pooled PPV was highest for self-harm plus suicide ($k = 13$) 38.9%, followed by self-harm alone ($k = 56$) 27.5% and suicide ($k = 35$) 3.7%. For the biological measures, for any outcome ($k = 24$) the pooled PPV was 15.0%, for self-harm ($k = 6$) 14.7% and suicide ($k = 16$) 14.5%. For the third-generation scales ($k = 19$) the pooled PPV for self-harm or self-harm plus suicide was 38.7%; for general hospital populations ($k = 46$) it was 32.5% and for psychiatric hospital in-patients ($k = 15$) it was 26.8% (Fig. 2).

For the individual biological tests predicting suicide, the best pooled PPV was for CSF 5-HIAA ($k = 6$) 21.1%; for individual psychological scales predicting self-harm or self-harm plus suicide combined, the BHS ($k = 4$) 29.1% or the Buglass and Horton scale ($k = 9$) 28.8% were equal; and for third-generation scales the ERRS ($k = 3$) 27.6% was best (Fig. 3).

Heterogeneity and risk of publication bias

The I^2 statistics (Figs 2 and 3) indicated a high degree of heterogeneity among scales, except for the CSF 5-HIAA. The funnel plots using all $k = 128$ scales for all outcomes also suggested heterogeneity is present. For scales with large sample sizes (Fig. 4, top half of the plot), the scale-specific PPVs fall evenly on either side of the pooled PPV (on the logit scale). However, for scales with smaller sample sizes (bottom half of the plot), more studies appear to have been published with high PPV values. However, it is unclear whether the pattern is indicative of heterogeneity or publication bias.[90]
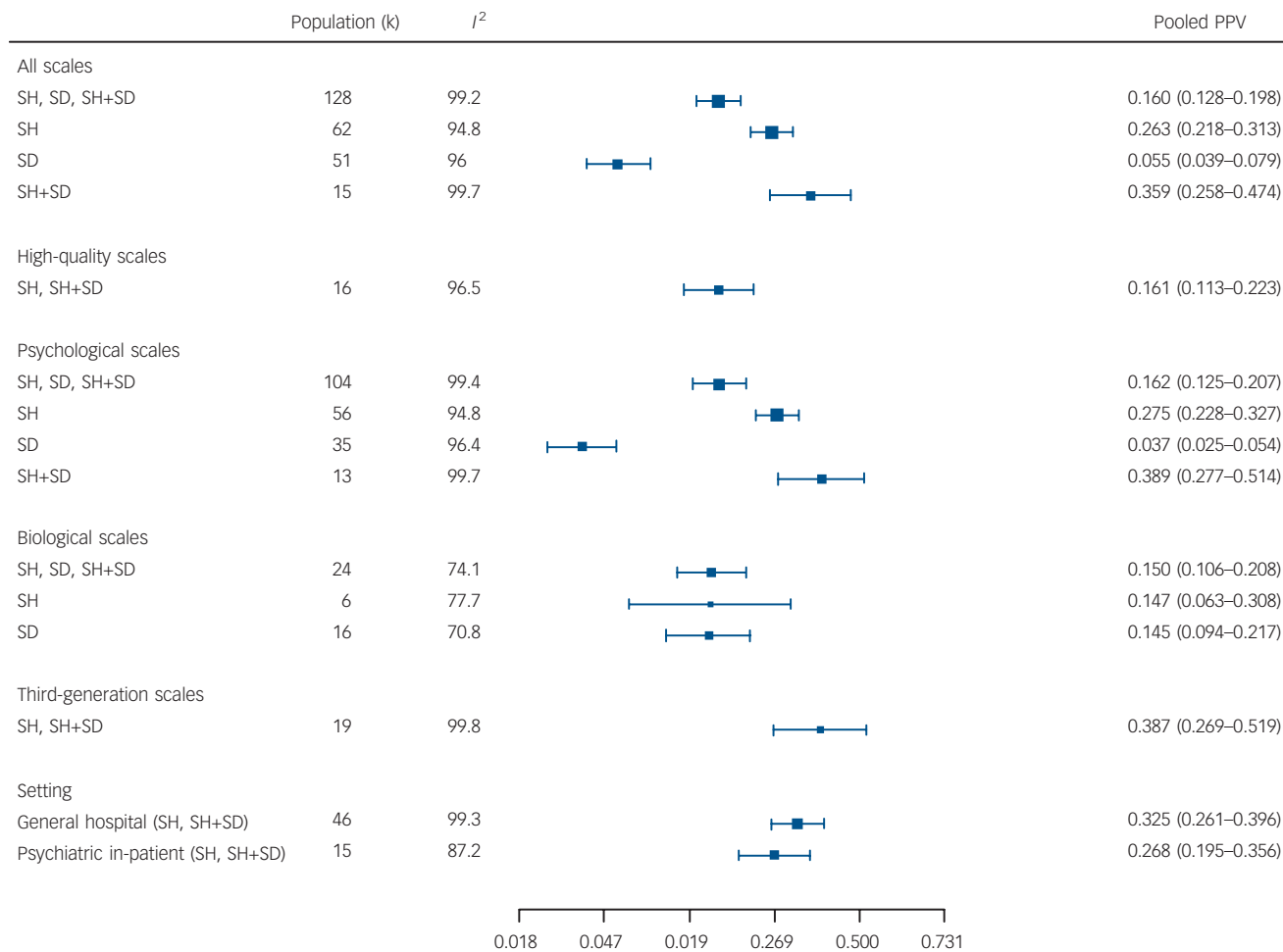
	Population (k)	I^2		Pooled PPV
All scales				
SH, SD, SH+SD	128	99.2		0.160 (0.128–0.198)
SH	62	94.8		0.263 (0.218–0.313)
SD	51	96		0.055 (0.039–0.079)
SH+SD	15	99.7		0.359 (0.258–0.474)
High-quality scales				
SH, SH+SD	16	96.5		0.161 (0.113–0.223)
Psychological scales				
SH, SD, SH+SD	104	99.4		0.162 (0.125–0.207)
SH	56	94.8		0.275 (0.228–0.327)
SD	35	96.4		0.037 (0.025–0.054)
SH+SD	13	99.7		0.389 (0.277–0.514)
Biological scales				
SH, SD, SH+SD	24	74.1		0.150 (0.106–0.208)
SH	6	77.7		0.147 (0.063–0.308)
SD	16	70.8		0.145 (0.094–0.217)
Third-generation scales				
SH, SH+SD	19	99.8		0.387 (0.269–0.519)
Setting				
General hospital (SH, SH+SD)	46	99.3		0.325 (0.261–0.396)
Psychiatric in-patient (SH, SH+SD)	15	87.2		0.268 (0.195–0.356)

0.018   0.047   0.019   0.269   0.500   0.731

Fig. 2 Summary pooled positive predictive values (PPVs) from meta-analyses of all scales, psychological, biological, high-quality, third-generation scales and general hospital and psychiatric in-patient settings.
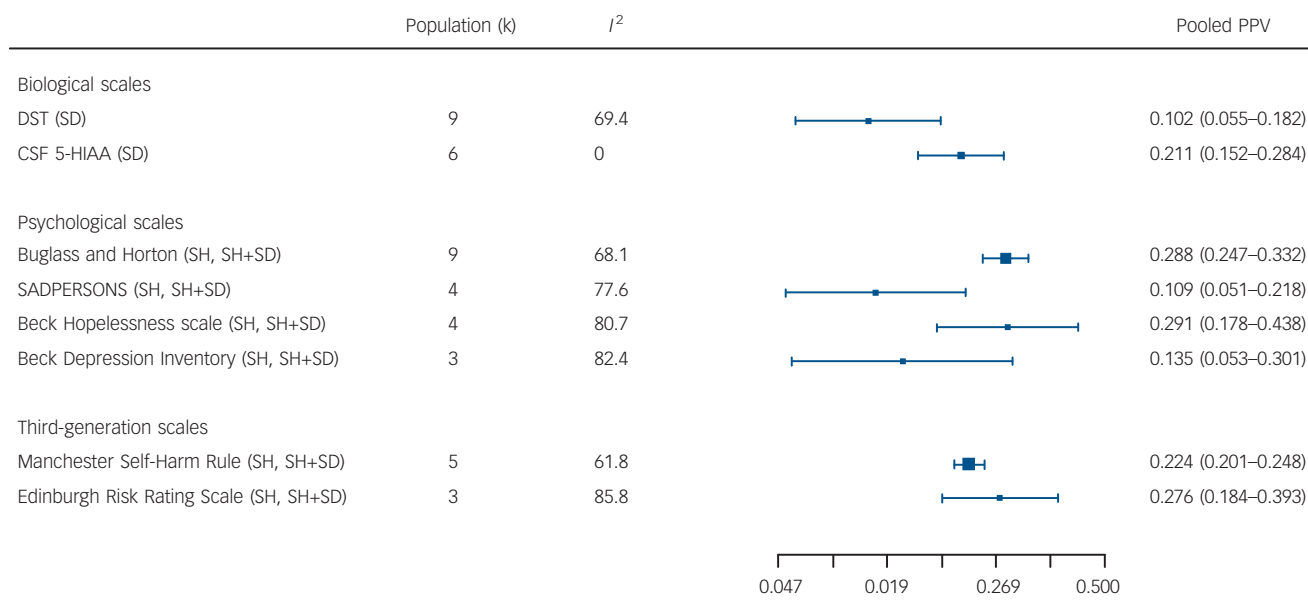
SD, suicide death, SH, self-harm.

	Population (k)	I^2		Pooled PPV
Biological scales				
DST (SD)	9	69.4		0.102 (0.055–0.182)
CSF 5-HIAA (SD)	6	0		0.211 (0.152–0.284)
Psychological scales				
Buglass and Horton (SH, SH+SD)	9	68.1		0.288 (0.247–0.332)
SADPERSONS (SH, SH+SD)	4	77.6		0.109 (0.051–0.218)
Beck Hopelessness scale (SH, SH+SD)	4	80.7		0.291 (0.178–0.438)
Beck Depression Inventory (SH, SH+SD)	3	82.4		0.135 (0.053–0.301)
Third-generation scales				
Manchester Self-Harm Rule (SH, SH+SD)	5	61.8		0.224 (0.201–0.248)
Edinburgh Risk Rating Scale (SH, SH+SD)	3	85.8		0.276 (0.184–0.393)

0.047   0.019   0.269   0.500

Fig. 3 Summary pooled positive predictive values (PPVs) from meta-analyses of specific biological scales, psychological scales and third-generation scales.

DST, Dexamethasone Suppression Test; CSF 5-HIAA, cerebrospinal fluid 5-hydroxyindoleacetic acid; SD, suicide death, SH, self-harm.
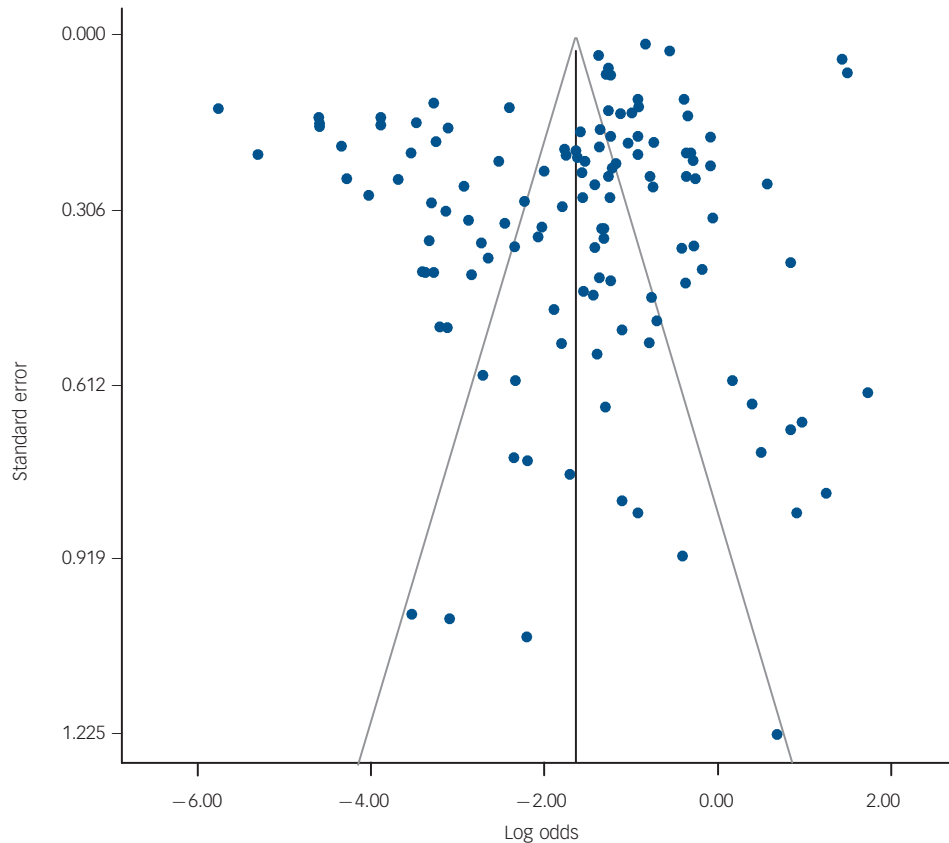
Fig. 4 Funnel plot for all scales and studies where the effect size of interest is positive predictive value (PPV).

Discussion

Prevalence rates and accuracy statistics

The PPV and NPV of all predictive instruments is limited by the prevalence of the outcome (i.e. 'disease') in the population of interest. This has been recognised in the prediction of suicide for over 60 years: 'Suicide is an infrequent event and its prediction is subject to the limitations found in the prediction of any infrequent behavior or event'.[91] To illustrate, Pokorny presented a theoretical calculation, using a prevalence of suicide of 500/100 000/year (the suicide rate for psychiatric in-patients in Pokorny's unit) combined with a hypothetical predictive test having 99% sensitivity and 99% specificity.[60] Under these idealised conditions the PPV was a modest 33%; and since 66% of positives would be false positives, the classification as 'high risk' was not useful to allocate intrusive and expensive treatment such as (involuntary) admission to hospital to prevent future suicide.[60] Pokorny suggested that a test with a more realistic 50% sensitivity and 90% specificity would yield a PPV of only 2%,[60] which is close to the pooled PPV estimates (range: 4–21%) from the current study.

Since repetition of hospital-treated self-harm has higher prevalence, could this be more suitable for a risk classification approach? In our study, the high-quality studies yielded a pooled estimate for PPV of 16.1% (including self-harm plus suicide), which is no different to the pooled prevalence estimate found by Carroll and colleagues.[92] The third-generation scales, most of which had a high risk of bias (inclined to maximise prevalence), had a pooled PPV of 38.7%, which appears to be an improvement over the pre-test probability of 16.3%. Could this be clinically useful? We address this question below.

Clinical utility of predictive tests

There are three methods used to determine the clinical utility of a predictive instrument: the PPV, the likelihood ratio (positive) (LR+)[14] and the clinical utility index (positive) (CUI+).[93] Similarly, there are three approaches to the question 'what are the best ways to decide whether my patient does not need treatment (or is safe to send home) based on a negative predictive test (i.e. classification as low risk)?', however, that question will need to be addressed in a separate study.

PPVS

The simplest is the PPV, the proportion of test-positive patients that will have the outcome; the balance being false positives. The clinician considers the available interventions, including efficacy, adverse effects and cost of administration and then makes a balanced judgement as to the usefulness of the positive test to allocate treatment. Involuntary admission to psychiatric hospital (to prevent suicide), which is highly intrusive, high cost, of unclear efficacy and with adverse effects on social standing, employment and health insurance status, would generally require a very high PPV to be considered useful. Conversely, an intervention which is effective (to prevent self-harm), brief, medium cost, delivered in the community, with low likelihood of adverse effects,[94] when balanced against the false positive patients receiving a treatment they did not need but was unlikely to harm them, might require a lower PPV.

Pre-test probabilities, post-test probabilities, likelihood ratios and Fagan nomograms

Likelihood ratios are said to be independent of the underlying prevalence rate, while being applicable to an individual.[14]

Likelihood ratios that are close to 1.0 have no clinical usefulness and a LR+ of more than 10 is likely to be clinically useful. Likelihood ratios can be calculated, LR+ = sensitivity/(1−specificity), but in practice it is easier to use the Fagan's nomogram that graphically links pre-test probability, likelihood ratio and post-test probability. Online versions of these nomograms are freely available (for example http://araw.mede.uic.edu/cgi-bin/testcalc.pl).

Taking the repetition rate of hospital-treated self-harm (16.3% in 12 months)[92] as the pre-test probability for any patient; the best case post-test probability was 29% for the Buglass and Horton Scale and the BHS (LR+ 2.1); and 39% for the third-generation scales (LR+ 3.3). Similarly for an in-patient in a psychiatric hospital (expected 6.5% self-harm in 12 months);[95] the pooled estimate of 27% as the post-test probability (LR+ of 5.2), would appear to be possibly useful. However, these study populations actually had a mean prevalence of 17.9% (LR+ 1.68), in which case prediction would have little usefulness.

CUI+

The CUI+ = sensitivity × PPV and is graded for utility: excellent ≥0.81, good ≥0.64, satisfactory ≥0.49 and poor <0.49.[93] Even when using the strongest results, the CUI+ was of poor utility: all scales, PPV 35.9%, pooled sensitivity 67.3% (CUI+0.24); psychological scales, PPV 38.9%, pooled sensitivity 70.0% (CUI+0.27); and third-generation studies, PPV 38.7%, pooled sensitivity 84.0% (CUI+0.33).

Duration of follow-up and clinical assessment of future suicidal behaviour

We considered 12 months as the longest duration of meaningful follow-up for clinical relevance and service organisation planning. Randomised controlled trials of psychosocial interventions are usually evaluated over a period of 6 or 12 months for the repetition of self-harm outcome.[96] Many of the primary studies identified in our review used much longer follow-up, with a resulting increased prevalence rate of the outcomes and hence improved PPV estimates. This can be seen in the biological scales predicting suicide; the best pooled PPV was for CSF 5-HIAA ($k = 6$) 21.1%. This result was strongly influenced by six studies[73,74,82,83,85,89] where the sample sizes were small, and the populations were psychiatric in-patients (mostly with a depression diagnosis). The risk of bias was high or unclear for patient selection and the follow-up period was longer than 12 months for five evaluations. The prevalence of suicide in these six studies ranged from 3 to 33% (mean 17%), which is many times the expected rate for unselected psychiatric in-patients of 0.5% at 12 months after discharge;[97] and more similar to a 19% lifetime prevalence for in-patient-treated populations with depression.[98]

Can risk assessment be used in clinical practice to determine allocation of intervention?

Our meta-analysis shows that no instrument is sufficiently accurate as a basis to determine allocation to intervention. We would not recommend that 'risk assessment' be used to classify patients in order to allocate follow-up care, since most patients will be incorrectly classified (false positives) and directed to unnecessary treatment, whereas many patients will be classified as low risk (false negatives) and hence be denied necessary treatment. This is consistent with the NICE Clinical Guideline 133, which suggests that scales should not be used to predict future suicide or repetition of self-harm[13] and a recent review focused on a small number of predictive instruments.[17]

Alternatives to the risk assessment stratification approach to treatment allocation

Perhaps, the notion of 'comprehensive risk assessment' can be integrated into clinical practice with 'comprehensive clinical assessment',[12] without the need to stratify patients into highly inaccurate risk categories. We would suggest that there are at least three alternative approaches to help determine treatment allocation.

First, clinical assessment can be used to identify any modifiable risk factors with follow-up care allocated to reduce exposure to those risks. Examples include: evidence-based treatments (for example for mood, substance use, psychotic or borderline personality disorders) or clinically accepted treatments (for example for relationship problems) or accepted standards of general care (for example individual and family support, social involvement, financial support, restriction of access to means). This approach is consistent with the 'needs-based approach' advocated by NICE[13] and with a public health approach that seeks to reduce exposure to known modifiable risk factors, in order to reduce prevalence and incidence of suicidal behaviours. This approach can be used for hospital-treated self-harm and for psychiatric in-patients at the time of discharge. Second, in sub-populations of patients who self-harm, for example patients meeting criteria for borderline personality disorder, there is proven efficacy for several psychological interventions specifically to reduce the number of self-harm events,[99] and these interventions are probably underutilised in clinical practice. Third, in unselected hospital-treated self-harm populations, cognitive–behavioural-based psychotherapy interventions have proven efficacy to reduce the proportion with any future self-harm;[96,100] and brief contact interventions may reduce the number of self-harm events.[101] Patients who have self-harmed who are hospital treated could be allocated to these effective treatments without risk stratification. However, since 84% of patients will not repeat self-harm in 12 months, low-cost, short-term treatments with fewer adverse effects should be given higher priority. NICE guidelines suggest 'Consider offering 3 to 12 sessions of a psychological intervention that is specifically structured for people who self-harm'.[13] Much less is known about interventions for the psychiatric in-patient population following discharge and these populations merit the development and evaluation of interventions to reduce subsequent self-harm.

Practice and policy implications

No individual predictive instrument or pooled subgroups of instruments were able to classify patients as being at high risk of suicidal behaviour with a level of accuracy suitable to be used to allocate treatment. Low prevalence outcomes, i.e. suicidal behaviours, are unlikely to be predicted by any instrument, even in key high-risk clinical populations, because of the statistical relationship of prevalence to PPV. The fairly steady increase in publication of papers arguing for the benefits of various risk assessment instruments and the parallel recommendations of prominent suicide prevention bodies to embrace the risk stratification approach for allocation of interventions has persisted despite the evidence against the clinical usefulness of this approach. Perhaps the evidence from this systematic review and meta-analysis will be used to mitigate these phenomena.

We would recommend three alternative approaches to a risk-based assessment to allocate intervention for high-risk clinical populations: first, an individual needs-based assessment followed by intervention to meet patient needs and to reduce exposure to modifiable risk factors; second, allocation of proven interventions for particular subpopulations; and third, the allocation of proven

interventions that can be delivered to unselected clinical populations.

Limitations of the study

In any systematic review there is the danger of missing published studies because of incorrect selection of search terms or exclusion of studies based on assessment of titles and abstracts. Observational studies reporting accuracy of predictive instruments may be more difficult to identify than studies of recent randomised controlled trials, for which there are more established standards for titles and keywords. The risk of bias was high in some studies, particularly so for the studies of biological scales, which usually were much older studies and often capitalised on highly biased selection of participants and long follow-up times. Most studies of psychological scales examined hospital-treated self-harm populations and most biological tests were applied to in-patients in psychiatric hospitals with severe mood disorder, so generalisation of these findings to other populations should be done with caution. The meta-analysis of predictive studies differs from the meta-analysis of intervention studies in that heterogeneity is to be expected and hierarchical random-effects models are needed to estimate effect sizes.[102] There was a high degree of heterogeneity for PPVs and in part this must be attributed to the differences in prevalence for three main outcomes: suicide, self-harm and self-harm plus suicide. These variations in prevalence can be seen in Tables DS1 and DS2. The I^2 statistic overestimates heterogeneity in meta-analyses of diagnostic tests.[103] The further exploration of heterogeneity will require a series of meta-regressions,[103] which could not be done in the current paper because of space restrictions. There will be other sources of heterogeneity, particularly arising from the populations selected, the predictor variables, the measurement of outcomes and the period of follow-up, which will be investigated further in a future paper.

Gregory Carter, MBBS, Cert Child Psych, PhD, FRANZCP, Centre for Brain and Mental Health Research, University of Newcastle, New South Wales, Australia; **Allison Milner**, BJPsych (Hons), MEpi, PhD, Population Health Strategic Research Centre, Deakin University, Burwood, and Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Victoria, Australia; **Katie McGill**, MPsych (Clin), DClinPsych, Centre for Brain and Mental Health Research, University of Newcastle, New South Wales, Australia; **Jane Pirkis**, MPsych, MAppEpid, PhD, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Victoria, Australia; **Nav Kapur**, MBChB, MMedSci, MD, FRCPsych, Centre for Suicide Prevention, Manchester Academic Health Science Centre, University of Manchester, and Greater Manchester Mental Health NHS Foundation Trust, Manchester, UK; **Matthew J. Spittal**, MBiostat, PhD, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Victoria, Australia

Correspondence: Gregory Carter, Locked Bag #7, Hunter Region Mail Centre, 2301 NSW, Australia. Email: Gregory.carter@newcastle.edu.au

First received 7 Feb 2016, final revision 16 Oct 2016, accepted 16 Nov 2016

Funding

References

1 Meehan J, Kapur N, Hunt IM, Turnbull P, Robinson J, Bickley H, et al. Suicide in mental health in-patients and within 3 months of discharge. *Br J Psychiatry* 2006; **188**: 129–34.

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

3 Berman AL, Silverman MM. Suicide risk assessment and risk formulation. Part II: Suicide risk formulation and the determination of levels of risk. *Suicide Life Threat Behav* 2014; **44**: 432–43.

4 Beck AT, Ward CH, Mendelson MM, Mock JJ, Erbaugh JJ. An inventory for measuring depression. *Arch Gen Psychiatry* 1961; **4**: 561–71.

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

6 Carroll BJ. The dexamethasone suppression test for melancholia. *Br J Psychiatry* 1982; **140**: 292–304.

7 Carroll BJ, Greden JF, Feinberg M. Suicide, neuroendocrine dysfunction and CSF 5-HIAA concentrations in depression. *Recent Advances in Neuropsychopharmacology* (eds B Angrist, GD Burrows, M Lader, O Lingjaerde, G Sedvall, D Wheatley): 307–13. Pergamon, 1981.

8 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.

9 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.

10 Claassen CA, Harvilchuck-Laurenson JD, Fawcett J. Prognostic models to detect and monitor the near-term risk of suicide: state of the science. *Am J Prevent Med* 2014; **47** (suppl 2): S181–5.

11 Office of the Surgeon General and National Action Alliance for Suicide Prevention. *012 National Strategy for Suicide Prevention: Goals and Objectives for Action*. Department of Health and Human Services (HHS), 2012.

12 Ryan CJ, Large MM. Suicide risk assessment: where are we now? *Med J Australia* 2013; **198**: 462–3.

13 National Institute for Health and Care Excellence. *NICE Guidelines Self-Harm: Longer-Term Management (CG 133)*. NICE, 2011.

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

15 Freedenthal S. Assessing the wish to die: a 30-year review of the Suicide Intent Scale. *Arch Suicide Res* 2008; **12**: 277–98.

16 Warden S, Spiwak R, Sareen J, Bolton JM. The SAD PERSONS scale for suicide risk assessment: a systematic review. *Arch Suicide Res* 2014; **18**: 313–26.

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

18 Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; **339**: b2700.

19 Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; **155**: 529–36.

20 Stijnen T, Hamza TH, Ozdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Stat Med* 2010; **29**: 3046–67.

21 Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Statist Softw* 2010; **36**: 1–48.

22 R Core Team. *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, 2015.

23 Beck AT, Steer RA, Kovacs M, Garrison B. Hopelessness and eventual suicide: a 10-year prospective study of patients hospitalized with suicidal ideation. *Am J Psychiatry* 1985; **142**: 559–63.

24 Beck AT, Brown G, Steer RA. Prediction of eventual suicide in psychiatric inpatients by clinical ratings of hopelessness. *J Consult Clin Psychol* 1989; **57**: 309–10.

25 Beck AT, Brown G, Berchick RJ, Stewart BL, Steer RA. Relationship between hopelessness and ultimate suicide: a replication with psychiatric outpatients. *Am J Psychiatry* 1990; **147**: 190–5.

26 Beck AT, Brown GK, Steer RA, Dahlsgaard KK, Grisham JR. Suicide ideation at its worst point: a predictor of eventual suicide in psychiatric outpatients. *Suicide Life Threat Behav* 1999; **29**: 1–9.

27 Bilen K, Ponzer S, Ottosson C, Castren 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.

28 Bilen K, Ponzer S, Ottosson C, Castren M, Pettersson H. Deliberate self-harm patients in the emergency department: who will repeat and who will not? Validation and development of clinical decision rules. *Emerg Med J* 2013; **30**: 650–6.

29 Bolton JM, Spiwak R, Sareen J. Predicting suicide attempts with the SAD PERSONS scale: a longitudinal analysis. *J Clin Psychiatry* 2012; **73**: e735–41.

30 Brown GK, Beck AT, Steer RA, Grisham JR. Risk factors for suicide in psychiatric outpatients: a 20-year prospective study. *J Consult Clin Psychology* 2000; **68**: 371–7.

31 Buglass D, McCulloch JW. Further suicidal behaviour: the development and validation of predictive scales. *Br J Psychiatry* 1970; **116**: 483–91.

32 Buglass D, Horton J. A scale for predicting subsequent suicidal behaviour. *Br J Psychiatry* 1974; **124**: 573–8.

33 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.

34 Cohen E, Motto JA, Seiden RH. An instrument for evaluating suicide potential: a preliminary study. *Am J Psychiatry* 1966; **122**: 886–91.

35 Colman I, Newman SC, Schopflocher D, Bland RC, Dyck RJ. A multivariate study of predictors of repeat parasuicide. *Acta Psychiatr Scand* 2004; **109**: 306–12.

36 Corcoran P, Kelleher MJ, Kelley HS, Byrne S, Burke U, Williamson E. A preliminary statistical model for identifying repeaters of parasuicide. *Arch Suicide Res* 1997; **3**: 65–74.

37 Erdman HP, Greist JH, Gustafson DH, Taves JE, Klein MH. Suicide risk prediction by computer interview: a prospective study. *J Clin Psychiatry* 1987; **48**: 464–7.

38 Garzotto N, Siani R, Tansella CZ, Tansella M. Cross-validation of a predictive scale for subsequent suicidal behaviour in an Italian sample. *Br J Psychiatry* 1976; **128**: 137–40.

39 Harriss L, Hawton K, Zahl D. Value of measuring suicidal intent in the assessment of people attending hospital following self-poisoning or self-injury. *Br J Psychiatry* 2005; **186**: 60–6.

40 Hartl TL, Rosen C, Drescher K, Lee TT, Gusman F. Predicting high-risk behaviors in veterans with posttraumatic stress disorder. *J Nerv Ment Dis* 2005; **193**: 464–72.

41 Hawton K, Fagg J. Repetition of attempted suicide: the performance of the Edinburgh predictive scales in patients in Oxford. *Arch Suicide Res* 1995; **1**: 261–72.

42 Hendin H, Al Jurdi RK, Houck PR, Hughes S, Turner JB. Role of intense affects in predicting short-term risk for suicidal behavior: a prospective study. *J Nerv Ment Dis* 2010; **198**: 220–5.

43 Huth-Bocks AC, Kerr DCR, Ivey AZ, Kramer AC, King CA. Assessment of psychiatrically hospitalized suicidal adolescents: self-report instruments as predictors of suicidal thoughts and behavior. *J Am Acad Child Adolesc Psychiatry* 2007; **46**: 387–95.

44 Klonsky ED, Kotov R, Bakst S, Rabinowitz J, Bromet EJ. Hopelessness as a predictor of attempted suicide among first admission patients with psychosis: a 10-year cohort study. *Suicide Life Threat Behav* 2012; **42**: 1–10.

45 Kreitman N, Foster J. The construction and selection of predictive scales, with special reference to parasuicide. *Br J Psychiatry* 1991; **159**: 185–92.

46 Kurz A, Möller HJ, Torhorst A, Lauter H. Validation of six risk scales for suicide attempters. In *Current Issues of Suicidology* (eds HJ Möller, A Schmidtke, R Welz): 174–8. Springer, 1988.

47 Larzelere RE, Smith GL, Batenhorst LM, Kelly DB. Predictive validity of the Suicide Probability Scale among adolescents in group home treatment. *J Am Acad Child Adolesc Psychiatry* 1996; **35**: 166–72.

48 Motto JA, Heilbron DC. Development and validation of scales for estimation of suicide risk. In *Suicidology: Contemporary Developments* (ed. ES Schneidman): 169–99. Grune & Stratton, 1976.

49 Motto JA, Heilbron DC, Juster RP. Development of a clinical instrument to estimate suicide risk. *Am J Psychiatry* 1985; **142**: 680–6.

50 Myers ED. Predicting repetition of deliberate self-harm: a review of the literature in the light of a current study. *Acta Psychiatr Scand* 1988; **77**: 314–9.

51 Naud H, Daigle MS. Predictive validity of the Suicide Probability Scale in a male inmate population. *J Psychopathol Behav Assess* 2010; **32**: 333–42.

52 Nimeus A, Traskman-Bendz L, Alsen M. Hopelessness and suicidal behavior. *J Affect Disord* 1997; **42**: 137–44.

53 Nimeus A, Alsen M, Traskman-Bendz L. The Suicide Assessment Scale: an instrument assessing suicide risk of suicide attempters. *Eur Psychiatry* 2000; **15**: 416–23.

54 Nimeus A, Alsen M, Traskman-Bendz L. High suicidal intent scores indicate future suicide. *Arch Suicide Res* 2002; **6**: 211–9.

55 Nock MK, Park JM, Finn CT, Deliberto TL, Dour HJ, Banaji MR. Measuring the suicidal mind: implicit cognition predicts suicidal behavior. *Psychol Sci* 2010; **21**: 511–7.

56 Ojehagen A, Danielsson M, Traskman-Bendz L. Deliberate self-poisoning: treatment follow-up of repeaters and nonrepeaters. *Acta Psychiatr Scand* 1992; **85**: 370–5.

57 Pallis DJ, Gibbons JS, Pierce DW. Estimating suicide risk among attempted suicides: II. Efficiency of predictive scales after the attempt. *Br J Psychiatry* 1984; **144**: 139–48.

58 Perry AE, Gilbody S. Detecting and predicting self-harm behaviour in prisoners: a prospective psychometric analysis of three instruments. *Soc Psychiatry Psychiatr Epidemiol* 2009; **44**: 853–61.

59 Petrie K, Chamberlain K, Clarke D. Psychological predictors of future suicidal behaviour in hospitalized suicide attempters. *Br J Clin Psychol* 1988; **27**: 247–57.

60 Pokorny AD. Prediction of suicide in psychiatric patients: report of a prospective study. *Arch Gen Psychiatry* 1983; **40**: 249–57.

61 Randall JR, Rowe BH, Colman I. Emergency department assessment of self-harm risk using psychometric questionaries. *Can J Psychiatry* 2012; **57**: 21–8.

62 Randall JR, Rowe BH, Dong KA, Nock MK, Colman I. Assessment of self-harm risk using implicit thoughts. *Psychol Assess* 2013; **25**: 714–21.

63 Roaldset JO, Linaker OM, Bjorkly S. Predictive validity of the MINI Suicidal Scale for self-harm in acute psychiatry: a prospective study of the first year after discharge. *Arch Suicide Res* 2012; **6**: 287–302.

64 Sanchez-Gistau V, Baeza I, Arango C, Gonzalez-Pinto A, de la Serna E, Parellada M, et al. Predictors of suicide attempt in early-onset, first-episode psychoses: a longitudinal 24-month follow-up study. *J Clin Psychiatry* 2013; **74**: 61–8.

65 Sapyta J, Goldston DB, Erkanli A, Daniel SS, Heilbron N, Mayfield, A et al. Evaluating the predictive validity of suicidal intent and medical lethality in youth. *J Consult Clin Psychology* 2012; **80**: 222–31.

66 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* 2013; **31**: 796–8.

67 Siani R, Garzotto N, Zimmermann Tansella C, Tansella M. Predictive scales for parasuicide repetition: further results. *Acta Psychiatr Scand* 1979; **59**: 17–23.

68 Sidley GL, Calam R, Wells A, Hughes T, Whitaker K. The prediction of parasuicide repetition in a high-risk group. *Br J Clin Psychol* 1999; **38**: 375–86.

69 Steeg S, Kapur N, Webb R, Applegate E, Stewart SLK, 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.

70 Waern M, Sjostrom 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.

71 Yaseen ZS, Kopeykina I, Gutkovich Z, Bassirnia A, Cohen LJ, Galynker II. Predictive validity of the Suicide Trigger Scale (STS-3) for post-discharge suicide attempt in high-risk psychiatric inpatients. *PLoS One* 2014; **9**: e86768.

72 Yen S, Shea MT, Walsh Z, Edelen MO, Hopwood CJ, Markowitz JC, et al. Self-harm subscale of the Schedule for Nonadaptive and Adaptive Personality (SNAP): predicting suicide attempts over 8 years of follow-up. *J Clin Psychiatry* 2011; **72**: 1522–8.

73 Asberg M, Traskman L, Thoren P. 5-HIAA in the cerebrospinal fluid: a biochemical suicide predictor? *Arch Gen Psychiatry* 1976; **33**: 1193–7.

74 Asberg M, Nordstrom P, Traskman-Bendz L. Biological factors in suicide. In *Suicide* (ed. A Roy): 47–71. Williams and Wilkins, 1986.

75 Black DW, Monahan PO, Winokur G. The relationship between DST results and suicidal behavior. *Ann Clin Psychiatry* 2002; **14**: 83–8.

76 Boza RA, Milanes FJ, Llorente M, Reisch J, Slater VL, Garrigo L. The DST and suicide among depressed alcoholic patients. *Am J Psychiatry* 1988; **145**: 266–7.

77 Coryell W, Young E, Carroll B. Hyperactivity of the hypothalamic-pituitary-adrenal axis and mortality in major depressive disorder. *Psychiatry Res* 2006; **142**: 99–104.

78 De Leo D, Pellegrini C, Serraiotto L, Magni G, De Toni R. Assessment of severity of suicide attempts: a trial with the dexamethasone suppression test and 2 rating scales. *Psychopathol* 1986; **19**: 186–91.

79 Edman G, Asberg M, Levander S, Schalling D. Skin conductance habituation and cerebrospinal fluid 5-hydroxyindoleacetic acid in suicidal patients. *Arch Gen Psychiatry* 1986; **43**: 586–92.

80 Galfalvy H, Huang YY, Oquendo MA, Currier D, Mann JJ. Increased risk of suicide attempt in mood disorders and TPH1 genotype. *J Affect Disord* 2009; **115**: 331–8.

81 Jokinen J, Carlborg A, Martensson B, Forslund K, Nordstrom AL, Nordstrom P. DST non-suppression predicts suicide after attempted suicide. *Psychiatry Res* 2007; **150**: 297–303.

82 Jokinen J, Nordstrom AL, Nordstrom P. CSF 5-HIAA and DST non-suppression – orthogonal biologic risk factors for suicide in male mood disorder inpatients. *Psychiatry Res* 2009; **165**: 96–102.

83 Nordstrom P, Samuelsson M, Asberg M, Traskman-Bendz L, Aberg-Wistedt A, Nordin C, et al. CSF 5-HIAA predicts suicide risk after attempted suicide. *Suicide Life Threat Behav* 1994; **24**: 1–9.

84 Plocka-Lewandowska M, Araszkiewicz A, Rybakowski JK. Dexamethasone suppression test and suicide attempts in schizophrenic patients. *Eur Psychiatry* 2001; **16**: 428–31.

85 Roy A, Agren H, Pickar D, Linnoila M, Doran AR, Cutler NR, et al. Reduced CSF concentrations of homovanillic acid and homovanillic acid to 5-hydroxyindoleacetic acid ratios in depressed patients: relationship to suicidal behavior and dexamethasone nonsuppression. *Am J Psychiatry* 1986; **143**: 1539–45.

86 Roy A, de Jong J, Linnoila M. Cerebrospinal fluid monoamine metabolites and suicidal behavior in depressed patients: a 5-year follow-up study. *Arch Gen Psychiatry* 1989; **46**: 609–12.

87 Targum SD, Rosen L, Capodanno AE. The dexamethasone suppression test in suicidal patients with unipolar depression. *Am J Psychiatry* 1983; **140**: 877–9.

88 Yerevanian BI, Feusner JD, Koek RJ, Mintz J. The dexamethasone suppression test as a predictor of suicidal behavior in unipolar depression. *J Affect Dis* 2004; **83**: 103–8.

89 Samuelsson M, Jokinen J, Nordstrom A-L, Nordstrom P. CSF 5-HIAA, suicide intent and hopelessness in the prediction of early suicide in male high-risk suicide attempters. *Acta Psychiatr Scand* 2006; **113**: 44–7.

90 Lau J, Ioannidis JPA, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. *BMJ* 2006; **333**: 597–600.

91 Rosen A. Detection of suicidal patients: an example of some limitations in the prediction of infrequent events. *J Consult Psychology* 1954; **18**: 397–403.

92 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.

93 Mitchell AJ. Sensitivity x PPV is a recognized test called the clinical utility index (CUI+). *Eur J Epidemiol* 2011; **26**: 251–2.

94 Guthrie E, Kapur N, Mackway-Jones K, Chew-Graham C, Moorey J, Mendel E, et al. Randomised controlled trial of brief psychological intervention after deliberate self poisoning. *BMJ* 2001; **323**: 135–7.

95 Gunnell D, Hawton K, Ho D, Evans J, O'Connor S, Potokar J, et al. Hospital admissions for self harm after discharge from psychiatric inpatient care: cohort study. *BMJ* 2008; **337**: a2278.

96 Hetrick SE, Robinson J, Spittal MJ, Carter G. Effective psychological and psychosocial approaches to reduce repetition of self-harm: a systematic review, meta-analysis and meta-regression. *BMJ Open* 2016; **6**: e011024.

97 Goldacre M, Seagroatt V, Hawton K. Suicide after discharge from psychiatric inpatient care. *Lancet* 1993; **342**: 283–6.

98 Goodwin FK, Jamison KR. *Manic-Depressive Illness*. Oxford University Press, 1990.

99 National Health and Medical Research Council. *Clinical Practice Guideline for the Management of Borderline Personality Disorder*. National Health and Medical Research Council, 2012.

100 Hawton K, Witt KG, Taylor Salisbury TL, Arensman E, Gunnell D, Hazell P, et al. Psychosocial interventions for self-harm in adults. *Cochrane Database Syst Rev* 2016; **5**: CD012189.

101 Milner AJ, Carter G, Pirkis J, Robinson J, Spittal MJ. Letters, green cards, telephone calls and postcards: systematic and meta-analytic review of brief contact interventions for reducing self-harm, suicide attempts and suicide. *Br J Psychiatry* 2015; **206**: 184–90.

102 Macaskill P, Gatsonis C, Deek JJ, Harbord RM, Takwoingi Y. Analysing and presenting results. In *Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0* (eds JJ Deeks, PM Bossuyt, C Gatsonis): 4–59. Cochrane Collaboration, 2010.

103 Bossuyt P, Davenport C, Deeks J, Hyde C, Leeflang M, Scholten R. Interpreting results and drawing conclusions. In *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.9* (eds JJ Deeks, PM Bossuyt, C Gatsonis): 3–31. Cochrane Collaboration, 2013.

EXTRA
CONTENT
ONLINE