

Risk assessment scales to predict risk of hospital treated repeat self-harm: A cost-effectiveness modelling analysis

Author Accepted Manuscript

*Leah Quinlivan, PhD¹; *Sarah Steeg, PhD¹; Jamie Elvidge, PhD²; Rebecca Nowland, PhD¹; Prof Linda Davies, PhD²; Prof Keith Hawton, FMedSci³; Prof David Gunnell, DSc^{4,5}; Prof Nav Kapur, FRCPsych^{1,6}

1. Centre for Mental Health and Safety, University of Manchester, England
2. Division of Population Health, Health Services Research and Primary Care, University of Manchester, England
3. Centre for Suicide Research, University Department of Psychiatry, Warneford Hospital, Oxford, England
4. School of Social and Community Medicine, University of Bristol, England
5. National Institute for Health Research Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and University of Bristol
6. Greater Manchester Mental Health NHS Foundation Trust, Manchester, England

* Joint first authors

Corresponding author: Dr Leah Quinlivan

Address: Jean McFarlane Building, University of Manchester, Oxford Road, Manchester M13 9PL

Telephone: +44 (0)161 275 0727

Email: leah.quinlivan@manchester.ac.uk

This is the author accepted manuscript. It is advisable to refer to the publisher's version of the paper if you intend to cite from the work. The article has been accepted for publication in the Journal of Affective Disorders.

Keywords: self-harm, suicidal behaviour, emergency services, economic evaluation, risk scales

ABSTRACT

Background: Risk scales are used widely for assessing individuals presenting to Emergency Departments (EDs) following self-harm. There is growing evidence that risk scales have limited clinical utility in identifying episodes at highest risk of repeat self-harm. However, their cost-effectiveness in terms of treatment allocation and subsequent repeat self-harm is unknown. We aimed to examine the cost-effectiveness of five risk scales (SAD PERSONS Scale, Modified SAD PERSONS Scale, ReACT Self-Harm Rule, Manchester Self-Harm Rule, Barratt Impulsivity Scale) and single item clinician and patient ratings of risk.

Method: Quality-Adjusted Life Years were estimated for each episode. The five risk scales and the patient rating were compared to the clinician rating. Incremental cost-effectiveness ratios (ICERs) were estimated for each scale, using a range of ICER thresholds. Sensitivity analysis explored different model assumptions.

Results: The formal scales were less cost-effective than the clinician and patient ratings across a range of ICER thresholds (£0-£30,000). The five scales were also less cost-effective than the clinician rating in most alternative scenario analyses. However, the clinician rating would be likely to result in unnecessary treatment costs for over half of patients identified as high risk.

Limitations: Our primary model depended on the assumption that high-intensity care reduced patients' risk of further self-harm.

Conclusion: The use of formal assessment tools for managing self-harm presentations to EDs did not appear to be cost-effective. While the judgement of a mental health clinician was found to be slightly more cost-effective, it still resulted in incorrect allocation of costs and missed treatment opportunities.

1 INTRODUCTION

2
3 Self-harm (intentional self-poisoning or self-injury) is an important public health issue (World Health
4 Organization, 2014). Over 200,000 people present to emergency departments (EDs) with self-harm
5 each year in England (Hawton et al., 2007). One in twenty five of these people will die by suicide within
6 ten years and around 16% will present to hospital with a repeat episode within a year (Carroll et al.,
7 2014). The average cost per attendance to an ED following self-harm was recently estimated to be
8 £809 (Tsiachristas et al., 2017). However, there is evidence that healthcare costs associated with self-
9 harm are a small fraction of the wider economic costs, notwithstanding the personal costs to
10 individuals involved (McDaid et al., 2017).

11 Multiple economic, social and psychological difficulties are associated with self-harm (Barnes et al.,
12 2016; Haw and Hawton, 2008; Hawton et al., 2016a; Hume and Platt, 2007), including complications
13 due to psychiatric diagnoses (Haw et al., 2001; Haw et al., 2005; Mars et al., 2014), alcohol (Ness et
14 al., 2015), and stressful life events (Townsend et al., 2016). The medical care of people who self-harm
15 is often challenging for health services (Sinclair et al., 2011). Potential care strategies vary in intensity
16 and cost e.g., resource intensive inpatient care versus less costly, less intensive outpatient, community
17 or primary care (Sinclair et al., 2011).

18 In England, there are at least 20 risk scales commonly used as part of the assessments to determine
19 patient management and risk of future self-harm (Quinlivan et al., 2014a). Evidence of their
20 effectiveness is limited and recommendations about their use vary across clinical guidelines (NICE,
21 2011; Royal College of Psychiatrists, 2010). The predictive utility of risk scales for repeat self-harm and
22 suicide is poor (Bolton et al., 2015; Carter et al., 2017; Quinlivan et al., 2016; Quinlivan et al., 2017).
23 Clinician and patient estimations of risk have been found to outperform existing risk scales in
24 predicting repeat self-harm, including one of the most widely used scales (the SAD PERSONS scale)
25 (Quinlivan et al., 2017).

However, the evidence on the use of risk assessments as interventions is more mixed. Two randomised controlled trials have been carried out in forensic settings. One found that the use of a structured risk assessment in the first few days of inpatient admission was associated with reductions in the incidence of violent episodes (Abderhalden et al., 2008). Another trial including risk assessment as part of an intervention in outpatient settings found that the assessment was not associated with reductions in criminal or violent incidents (Troquete et al., 2013). There is limited experimental evidence about risk assessment as an intervention for people who present to hospital with self-harm. A recent study from the United States found that combining screening with a brief intervention following an ED presentation was associated with reductions in repeat suicide attempts and overall number of repeat attempts (Miller et al., 2017).

While studies to date have examined predictive properties of risk scales, there are few studies of their associated costs. It is not known if there are any conditions under which a particular scale might be cost-effective in terms of treatment allocation and subsequent reductions in repeat self-harm.

The objectives of this study were to develop an economic model synthesising clinical and cost-related evidence to explore:

1. The cost-effectiveness of different risk assessment scales from the perspective of the hospital service;
2. The potential cost-effectiveness of scales under different conditions.

METHODS

Before constructing the model, a literature search of the UK National Health Service, Economic Evaluation Database (<https://www.evidence.nhs.uk>), a robust collection of published economic evaluations, was conducted in August 2017 to identify existing modelling studies of self-harm risk assessments or screening. The search, for studies containing “self-harm” or “suicide” in the title,

abstract or keywords, returned 40 studies. None presented an economic model relevant to the present decision problem (25 included drug interventions, 8 focused on psychotherapy interventions; 2 were treatments for patients with schizophrenia, one was a bridge barrier, one was a workplace intervention, one was a helpline, one was not relevant to suicide or self-harm and one was a duplicate).

This study followed the Consolidated Health Economic Evaluation Reporting Standards guidance {Husereau, 2013 #984} (see Table S1 for the checklist).

Participants

The population of interest for this study were individuals aged 18 years and over attending an emergency department in England following self-harm.

Model structure

A data-driven economic model was developed through collaboration with an interdisciplinary team of health economists, psychiatrists, and clinical researchers (Figure 1). A probability tree mapping out the immediate consequences of using a risk assessment scale was developed. When a patient presented at the emergency department after an episode of self-harm, they each had a risk of future self-harm that was unclear at the time of assessment. A risk assessment gave either a positive result, predicting the patient as at high risk of repetition, or a negative result, suggesting the patient was low risk. In the model the branches, or events, correspond to initial true-positive (correctly identified as high risk of repeat self-harm), false-positive (incorrectly identified as high risk of repeat self-harm), true-negative (correctly identified as low risk of repeat self-harm) or false-negative (incorrectly

identified as low risk of repeat self-harm) assessments. Misclassification and the effect of subsequent care on the risk of repeat self-harm are incorporated in to the economic model separately.

Figure 1

Cost-effectiveness model structure

In the model, patients classed as high risk of future self-harm by the risk assessment scale would receive high-intensity care, for example admission to a psychiatric bed or referral to outpatient psychiatric treatment. In contrast, those patients identified by the risk scale as having low risk of future self-harm would receive low-intensity care, which was determined to be one general practitioner (GP) consultation.

Patients who repeated self-harm (i.e., true-positive or false-negative risk assessments) were assigned a probability of further self-harm at any given time, based on real-world repetition rates reported in observational cohort studies (Table S2). While some individuals could have repeated self-harm and received care without presenting to the ED, we were interested in hospital treated self-harm, therefore we assumed that no patients without repeat hospital presentations (i.e., true-negative and false-positive) repeated self-harm. Patients were able to move into different health states in one-week cycles. For example, patients who repeated self-harm (either true-positive or false-negative classifications) could self-harm again and move to the repeat episode state. When a patient repeated self-harm, they returned to the start of the model, receiving another risk assessment to predict the likelihood of future self-harm. Those who did not repeat self-harm remained in the 'no repeat' health

state, represented by the semi-circular arrows. Patients could die in any health state, at which point they were removed from the model.

In our primary analysis the model was run for three years. While model durations should be sufficient to capture all major differential outcomes, longer time horizons suffer from increased uncertainty (NICE, 2013). Furthermore, the highest risk of repeat self-harm is within the first six months of the index episode (Kapur et al., 2006). In order to reflect society's higher value placed on costs and benefits today than in the future (NICE, 2013), and in line with national guidance recommendations, model outcomes (QALYs and costs) were discounted (reduced) by 3.5% per year.

Clinical inputs

Clinical data used to populate the model were obtained from a prospective cohort study evaluating risk scales following hospital-treated self-harm in five hospitals in England. Full details of the study are reported elsewhere (Quinlivan et al., 2017). In short, that study involved 514 patients aged 18 years and over who were referred to liaison psychiatry service following self-harm and consented to take part in a multi-site prospective cohort study evaluating the predictive accuracy of risk scales following self-harm between March 2014 and January 2015. The liaison psychiatry services were provided by multi-disciplinary teams which included psychiatric nurses, social workers, consultant psychiatrists, and junior doctors. For detailed description of the services available, see our earlier study {Quinlivan, 2017 #860}. Patients were referred to liaison psychiatry from the ED, with assessments taking place either in the ED or on a medical ward.

The assessment scales used in the study were selected on the basis of a systematic review of the diagnostic accuracy of risk scales following self-harm (Quinlivan et al., 2014b) and a cohort study of their diagnostic accuracy (Quinlivan et al., 2017). Clinician and patient global ratings of risk of repeated

self-harm within six months (using a 1-10 Likert-type scale) were also included. Our economic model included: the SAD PERSONS Scale, Modified SAD PERSONS Scale, ReACT Self-Harm Rule, Manchester Self-Harm Rule, the Barratt Impulsivity Scale, and clinician and patient ratings.

Scale predictive accuracy

The predictive accuracy of each scale – sensitivity (correctly predicting repeat self-harm) and specificity (correctly predicting patients who did not repeat self-harm) – was obtained by comparing positive/negative assessment results with repeat self-harm episodes (Quinlivan et al., 2017).

Self-harm episodes

Clinical model inputs are provided in Supplementary Material (Table S2). A key model input was the underlying proportion of true “high-risk” individuals; that is, those who repeated self-harm and were identified as at high risk of repeating. In the cohort study 145 (30%) of 483 participants had at least one repeat episode during the six-month follow-up period (Quinlivan et al., 2017). This figure was used as the background risk of repetition among the model population. Patients who repeated self-harm were able to experience multiple further episodes, and this risk increased following each episode (Quinlivan et al., 2017).

Resource use and costs

We obtained service use data from the prospective cohort study previously described (Quinlivan et al., 2017) and from Department of Health statistics (Department of Health, 2015a, b, c). The types of resource use included in the model were grouped into either self-harm episode care or follow-up care (Table S2). All self-harm episodes were assumed to present to an emergency department, incurring

the unit cost of this attendance. A proportion of patients would have required an ambulance to reach hospital, and some would have been admitted to treat their injuries, incurring additional inpatient stay costs.

The model assumed follow-up care was determined by the risk assessment result. A classification of low risk of repetition caused the patient to be referred back to the care of their GP only. High-intensity care was provided following a high risk classification, resulting in a potential reduction in risk of repeat self-harm (see below, 'Risk reduction'). This was included in the model as several possible referrals, each incurring more health care resources than low-intensity care: inpatient psychiatric care, outpatient and community care, social worker support, and specialist drug and alcohol services. The likelihood of receiving each type of referral was informed by the rates of referrals in the previous cohort study, excluding those who received only a GP referral. All resource use and cost inputs are provided in Table S2.

Risk reduction

There are limited data on the impact of risk assessment results on subsequent clinical management. Therefore, in order to model the risk reduction potential of the scales in the economic model, we assumed a positive assessment (whether true or false) would lead to high-intensity care, while a negative assessment would lead to low-intensity care. The model included the following assumptions:

- Receiving appropriate high-intensity care prevented repetition in some patients who would otherwise have repeated self-harm;
- Low-intensity care did not prevent repetition in patients who would otherwise have repeated self-harm.

While there is RCT evidence examining the effect of specific psychological treatments (Hawton et al., 2016b), there is limited evidence quantifying the effect of different types of routine management on repeat self-harm in clinical settings. One study found psychosocial assessment could reduce the risk of repeat self-harm by 13-57% (Bergen et al., 2010) and another study suggested a 40% reduction, but only in two out of three of the centres studied (Kapur et al., 2013). Another found an odds ratio for repeat self-harm of 0.84 favouring psychological intervention over no intervention (Erlangsen et al., 2014). Based on that evidence, in our primary analysis we assumed that high-intensity care caused a 20% risk reduction. The importance of this value was explored in sensitivity analysis.

Time to repeat episode

Data from the prospective cohort study (Quinlivan et al., 2017) was also used to estimate the rate of repetition for high-risk patients. Exponential curves provided the best statistical fit to the data overall, compared to other types of parametric curve, based on the Akaike and Bayesian Information Criteria (Akaike, 1974; Schwarz, 1978). The resulting constant rate of self-harm was 13% per week for the first repeat episode, which increased with each further self-harm episode to 39% per week for people who had more than five repeat episodes.

Mortality

All-cause mortality outcomes were based on mortality rates from national UK mortality data (ONS, 2015). These values were adjusted to reflect higher mortality among self-harming individuals (standard mortality ratios were 4.1 for males and 3.2 for females) (Bergen et al., 2012). Mortality rates were also adjusted to reflect higher risk among those who were admitted to a medical bed following self-harm (hazard ratio 1.3) (Kapur et al., 2015).

1

2 **Health economic inputs**

3

4 **Quality of life**

5 Patient quality of life associated with repeated episodes of self-harm was included in the form of utility
6 weights, reflecting the value of a particular health state relative to full health (equal to 1), death (equal
7 to 0) and potentially worse than death (less than 0) (Devlin et al., 2017). Utility weights were combined
8 with the time spent experiencing them to calculate quality-adjusted life years (QALYs). For example, a
9 person living for two years with a utility value of 0.82 would accrue $(2 * 0.82) = 1.64$ QALYs.

10 We assumed that patients' quality of life at the start of the model, after the index self-harm episode,
11 was represented by a utility weight of 0.671. There is no specific utility weight for individuals who have
12 self-harmed so we used the value corresponding to utility weights for "mental/behavioural problems"
13 and "personal history of mental disorder" (Sullivan et al., 2011).

14 Limited evidence exists regarding quality of life during a self-harm episode. A study of 16 Dutch
15 clinicians (van Spijker et al., 2011) reported a utility decrement of 0.46 attributable to the mental
16 distress of a suicide attempt, similar to an older Australian study (Mathers et al., 2001). This was
17 converted to an approximate utility weight by subtracting it from 'perfect health': $(1 - 0.46) = 0.54$.
18 This utility prevailed during the model cycle in which a patient self-harmed and received emergency
19 care. Utility then recovered to its previous level, minus an assumed decrement of 0.02 per repeat
20 episode, to reflect the more severe clinical profile of people who continue to self-harm (Haw et al.,
21 2007).

22

23 **Synthesis of health and cost outcomes**

A 'reference' risk-assessment scale was selected, and its total cost and QALY results were compared to each other scale. The clinician global rating was used as the reference scale, because it corresponded best to usual treatment and might be conceptualised as a 'no risk assessment scale' or 'no intervention' strategy. For each comparison, the incremental cost-effectiveness ratio (ICER) was calculated as follows: $ICER = \text{Difference in total costs} / \text{Difference in total QALYs}$. ICERs lower than a given threshold value indicate the intervention is considered to be cost-effective. Until recently this was typically £20-30,000 of extra spending for each additional QALY gained (NICE, 2013). However, more recent empirical estimates suggest this value should be around £12,000 (Claxton et al., 2015). Therefore, we used a threshold value of £12,000, while also examining results using the higher ICER threshold of £20-30,000.

Uncertainty analysis

To characterise uncertainty in our model inputs, values were sampled from their underlying distributions, rather than relying on fixed, potentially imprecise mean values (Table S2). We sampled all input values 5,000 times for all analyses, then took the average result across those 5,000 unique "model runs". This allowed us to present a mean result and a measure of variance around that mean.

Additionally, we tested the sensitivity of our results to different model assumptions, exploring a number of alternative scenarios to account for the uncertainty around some of the conditions. These included reducing the underlying proportion of patients who repeated self-harm; assuming no increase in risk over time; changing the risk reduction associated with high-intensity care; altering the resources required by outpatient and community care referrals; using a higher ICER threshold and adjusting the analysis duration. This also allowed us to address our second objective: to see if the cost-effectiveness of the scales changed under different circumstances.

RESULTS

The Markov diagram shows the number of episodes in each risk category along with subsequent management based on the model parameters (Figure 2). The clinician rating classed 47% of the episodes as being at higher risk of repeat self-harm in the prospective cohort study. According to the model parameters in the present study, based on 100 self-harm episodes, 25 of these episodes would have received unnecessary high intensity care (Figure 2). Eight episodes that repeated were missed and would not have received high intensity care.

Figure 2

Cost-effectiveness results

The total QALYs per patient were 1.984 for the clinician rating (Table 1). The MSHR and the ReACT both produced a QALY gain compared to this reference scale (+0.002).

The SPS, the MSPS and the BIS all resulted in an average net loss of QALYs and increase in costs compared to the clinician rating. They avoided few repeat episodes as they led to only a small proportion of patients receiving effective high-intensity care, generating the fewest total QALYs. The clinician rating dominated (had lower costs and was at least as effective) as the SPS, the MSPS and the BIS, providing health gains for a lower total cost. The clinician rating also dominated the patient rating on average. The incremental costs of the patient rating were smaller than the other scales, reflected in its 22% likelihood of being cost-effective.

Figure 3. shows that no risk-assessment scale was likely to be cost-effective over the clinician rating across all ICER threshold values. At the upper end of the ICER threshold, higher sensitivity scales (the

1 Manchester Self-Harm Rule and the ReACT Self-Harm Rule) had a greater chance of being cost-
2 effective (around 40%).

3

4 *Figure 3*

Table 1
Model cost-effectiveness results

Assessment scale	QALYs**: Mean (SD) [95% CI]	QALYs**: \pm reference scale	Costs: \pm £860 (the reference scale)	ICER vs. reference scale***	Likelihood more cost-effective than clinician rating (£12,000 per QALY)
Clinician rating*	1.984 (0.139) [1.722, 2.273]	-	-		
Patient rating	1.983 (0.139) [1.721, 2.273]	0.000 [-0.002, 0.001]	£13 [£-20, £55]	Dominated by the clinician rating****	22%
ReACT	1.986 (0.139) [1.725, 2.274]	0.002 [0, 0.004]	£77 [£-6, £145]	£34,892	12%
MSHR	1.986 (0.139) [1.725, 2.274]	0.002 [0, 0.004]	£87 [£3, £157]	£38,083	9%
SPS	1.977 (0.140) [1.715, 2.269]	-0.007 [-0.012, 0]	£133 [£-65, £512]	Dominated by the clinician rating	4%
BIS	1.975 (0.140) [1.713, 2.268]	-0.008 [-0.015, 0]	£201 [£-79, £766]	Dominated by the clinician rating	4%
MSPS	1.979 (0.139) [1.717, 2.269]	-0.005 [-0.010, 0]	£101 [£-36, £367]	Dominated by the clinician rating	3%

Abbreviations: MSHR: Manchester Self-Harm Rule; ReACT: ReACT Self-Harm Rule; SPS: SAD PERSONS Scale; MSPS: Modified SAD PERSONS Scale

*Clinician rating is the reference scale; **QALYs: Quality-adjusted life years; ***ICER: incremental cost effectiveness ratio; ****Dominated: intervention is less costly and at least as effective as the comparison scale, i.e. the clinician rating scale (NICE guide to the methods of technology appraisal 2013)

1 **Sensitivity analysis**

2 When the underlying proportion of patients who repeated self-harm was reduced, or when the effect
3 of increasing risk due to continued self-harm was removed, the potential benefit of high-intensity care
4 was smaller and the BIS and the SPS became more cost-effective (Table 2).

5 If appropriate high-intensity care was presumed to be more effective – such that 50% of repeat
6 episodes were avoided – then scales with higher sensitivity, such as the Manchester Self-Harm Rule
7 and the ReACT Rule, became more cost-effective. However, the clinician rating remained the most
8 cost-effective on average. Alternatively, under the assumption that high-intensity care caused no
9 reduction in the risk of repeat self-harm, the Barratt Impulsivity Scale dominated all scales.

10 Assuming no negative quality of life impact associated with each self-harm episode made higher
11 sensitivity scales less cost-effective, because the benefit of preventing self-harm was smaller.

12 If referrals to outpatient and community care required more contacts per patient – rising from two in
13 the primary analysis to seven, which more closely matches clinical guidelines (NICE, 2011), but may be
14 less reflective of reality – the SAD PERSONS Scale and Barratt Impulsivity Scale, which would have led
15 to fewer contacts, became more cost-effective relative to the clinician rating (74-76% likely).

16

17 Shortening the model duration to six months, allowing less time for the small QALY benefits to be
18 realised, caused both the SAD PERSONS scale and Barratt Impulsivity Scale to be 86% likely to be more
19 cost-effective than the clinician rating. At 12 months and 60 months, the clinician rating became more
20 likely to be cost-effective than any of the scales.

Sensitivity analyses (the two scales most likely to be cost-effective vs. clinician rating are shown)	Assessment scale*	Incremental QALYs**: Mean [95 percentile limits]	Incremental Costs: Mean [95 percentile limits]	Mean ICER***	Likelihood more cost-effective than clinician rating
Underlying risk of any repeat self-harm reduced to 21% (Cooper et al., 2015)	BIS	-0.0032 [-0.0049, 0]	-£97 [-£166, -£23]	£30,580	93%
	SPS	-0.0026 [-0.0041, 0]	-£77 [-£133, -£17]	£29,857	92%
Underlying risk of self-harm repetition does not change after each self-harm event (fixed)	BIS	-0.0037 [-0.0009, 0.0004]	-£116 [-£191, -£38]	£31,668	95%
	SPS	-0.0030 [-0.0048, 0]	-£92 [-£152, -£29]	£30,819	94%
Increased effectiveness of high-intensity care (50% reduction in risk of self-harm)	ReACT	0.0034 [0, 0.0062]	£46 [-£32, £110]	£13,627	43%
	MSHR	0.0035 [0, 0.0064]	£59 [-£25, £118]	£15,386	38%
No reduction in risk due to high-intensity care	BIS	0 [0, 0]	-£362 [-£544, -£240]	Dominates the clinician rating	100%
	SPS	0 [0, 0]	£298 [-£454, -£195]	Dominates the clinician rating	100%
No utility decrement per episode	Patient rating	0 [-0.0003, 0.0001]	£13 [-£21, £56]	Dominated by clinician rating	23%
	SPS	-0.0009 [-0.0029, 0]	£139 [-£69, £535]	Dominated by clinician rating	13%
Outpatient and community referrals lead to 7 contacts, on average	SPS	-0.0066 [-0.0123, 0]	-£180 [-£457, £226]	£27,127	76%
	BIS	-0.0083 [-0.0153, 0]	-£217 [-£586, 365]	£26,065	74%
Time horizon (model duration) of 6 months	BIS	-0.0008 [-0.0018, 0]	-£71 [-£173, £53]	£84,125	86%
	SPS	-0.0007 [-0.0014, 0]	-£57 [-£138, £44]	£82,516	86%
Time horizon (model duration) of 12 months	SPS	-0.0019 [-0.0037, 0]	£30 [-£98, £229]	Dominated by clinician rating	28%
	BIS	-0.0024 [-0.0046, 0]	£41 [-£125, £299]	Dominated by clinician rating	27%
Time horizon (model duration) of 60 months	ReACT	0.0037 [0, 0.0072]	£77 [-£5, £144]	£20,659	23%
	Patient rating	-0.0007 [-0.0031, 0.0013]	£13 [-£21, £56]	Dominated by clinician rating	21%

No discounting of costs or benefits	Patient rating	-0.0004 [-0.0019, 0.0007]	£13 [-£21, £56]	Dominated by clinician rating	21%
	ReACT	0.0022 [0, 0.0042]	£76 [-£5, £144]	£34,596	11%

*Abbreviations: MSHR: Manchester Self-Harm Rule; ReACT: ReACT Self-Harm Rule; SPS: SAD PERSONS Scale; MSPS: Modified SAD PERSONS Scale; **ICER: incremental cost effectiveness; *** QUALYs: Quality-adjusted life years

Table 2 Sensitivity analyses

1 In order for the Manchester Self-Harm Rule to be cost effective at the threshold of £12,000 per
2 additional QALY gained, at its existing sensitivity of 97%, the estimated specificity would need to
3 increase from 20% to 36%. At this level, 45 out of 100 patients would receive unnecessary high
4 intensity treatment. At the £30,000 threshold, the estimated specificity would need to be 23%
5 resulting in 54 out of 100 patients receiving unnecessary treatment.

7 **DISCUSSION**

8 Existing studies of risk assessment tools have focussed on their psychometric properties, such as
9 sensitivity and positive predictive value for predicting repeat self-harm and future suicide. Economic
10 modelling provides estimates of costs associated with unnecessary treatment for individuals
11 incorrectly identified as high risk, and savings associated with effective treatment on quality of life.
12 This helps to place the use of risk tools in the wider context of the provision of treatment, providing
13 additional evidence about their effectiveness. Our study suggests that the clinician rating is a more
14 cost-effective use of clinical resources than the five risk assessment scales we tested. The clinician
15 rating was also among the most cost-effective approaches in the majority of alternative scenario
16 analyses we explored; it harboured a 'middle ground' between the extremes of high sensitivity and
17 high specificity of the scales. However, the clinician rating would be likely to result in unnecessary
18 treatment costs for over half of patients identified as high risk. Sensitivity analysis suggested it was
19 unlikely to be cost-effective if high-intensity care was ineffective at reducing self-harm, or when the
20 underlying proportion of people who repeat self-harm was small.

22 **Limitations**

23 This is the first economic model developed to assess the cost-effectiveness of self-harm risk
24 assessment scales in the context of wider clinical services. While existing studies have focused on the

accuracy of different scales in predicting repeat self-harm, the present study explores consequences of risk assessment in terms of health and cost outcomes. A number of scenarios were explored, helping to account for the uncertainty around the extent of follow-up care received by individuals and the effects on risks of future self-harm. Predictive accuracy inputs, and some of the resource use inputs, were taken directly from a recent multi-site prospective cohort study (Quinlivan et al., 2017).

Presenting reasonable simplifications of clinical reality is necessary of all economic models. In this model, high-risk patients, without appropriate management, were assumed to experience a future self-harm episode with certainty, whereas low-risk patients were assumed not to repeat self-harm. This was a necessary, data-driven assumption, since (i) a true-positive result would not have been true unless the patient had gone on to repeat self-harm, and (ii) it was necessary to link an assessment result, indirectly, to the subsequent risk of repetition.

To characterise uncertainty, a number of sensitivity analyses were tested, capturing major clinical, economic and modelling assumptions. All analyses were conducted probabilistically to reflect our uncertainty in the accuracy of model input values.

Our primary results depended on the assumption that high-intensity care would be effective at reducing a patient's risk of further self-harm. While this seems intuitively appropriate, the magnitude of effectiveness is unclear. Furthermore, in the prospective cohort study, clinicians conducted several assessment scales to each patient, meaning the results of any particular scale could not be linked to the care subsequently provided (Quinlivan et al., 2017). In addition, clinicians may have adapted subsequent treatment based on the battery of assessment results, potentially diluting the predictive accuracy values of the scales.

1 The prevalence of true positive results may influence the applicability of the findings to other
2 settings. The repetition rate in the Quinlivan et al. (2017) prospective cohort study was 30%,
3 indicating relatively high levels of underlying morbidity and need. People who present to emergency
4 departments and do not receive a mental health assessment or who self-harm in the community are
5 likely to have a lower repetition rate. This would result in lower predictive accuracy of the scales.

7 **Clinical and research implications**

8 In a previous study, the formal scales included here were found to have limited clinical utility in
9 predicting future self-harm, performing either no better, or worse, than the clinicians' ratings
10 (Quinlivan et al., 2017). The finding in the present study that they are also unlikely to be cost-effective
11 provides further evidence that they should not be used alone to determine subsequent treatment and
12 follow-up care. Even the most cost-effective approach reported in this study, the clinician rating, had
13 sub-optimal performance in terms of the unnecessary costs associated with treatment for those who
14 were incorrectly identified as high risk. Given the complexity of self-harm and its associations with a
15 wide range of health and social difficulties, future research could examine the cost-effectiveness of
16 treatment referrals in greater detail. For example, cost-effectiveness of referrals to drug and alcohol
17 services for people with substance misuse or alcohol problems presenting with self-harm could be
18 examined.

19
20 Further research into the effectiveness of different types of management would also be useful. We
21 considered effectiveness to be equal across all types of high-intensity referral, an umbrella term
22 capturing various types of referral over and above referral to a GP. However, it may be appropriate to
23 differentiate between these types of care in terms of effectiveness.

Conclusion

Previous studies have provided evidence of the predictive accuracy and psychometrics associated with different risk assessment scales to predict future self-harm. We have developed a cost-effectiveness model to estimate the health economic consequences of using different scales by exploring what comes next; that is, the types of clinical services used after a risk assessment. Our results suggested that the five formal risk scales were less cost-effective than the clinical judgement of a mental health clinician. Our analysis provides no evidence for the effectiveness of the use of any formal scale or clinician rating of risk.

Funding

This paper presents independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research Programme (Grant Reference Number RP-PG-0610-10026). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Figure 1: Figure 1: Cost-effectiveness model structure

Figure 2: Markov diagram showing the Clinician Global Rating (based on predictive ability performance in Quinlivan et al., 2017).

Figure 3: Cost-effectiveness acceptability curves for the risk scales

References

- Abderhalden, C., Needham, I., Dassen, T., Halfens, R., Haug, H.J., Fischer, J.E., 2008. Structured risk assessment and violence in acute psychiatric wards: randomised controlled trial. *British Journal of Psychiatry* 193, 44-50.
- Akaike, H., 1974. NEW LOOK AT STATISTICAL-MODEL IDENTIFICATION. *Ieee Transactions on Automatic Control* AC19, 716-723.
- Barnes, M.C., Gunnell, D., Davies, R., Hawton, K., Kapur, N., Potokar, J., Donovan, J.L., 2016. Understanding vulnerability to self-harm in times of economic hardship and austerity: a qualitative study. *Bmj Open* 6.
- Bergen, H., Hawton, K., Waters, K., Cooper, J., Kapur, N., 2010. Psychosocial assessment and repetition of self-harm: The significance of single and multiple repeat episode analyses. *Journal of Affective Disorders* 127, 257-265.
- Bergen, H., Hawton, K., Waters, K., Ness, J., Cooper, J., Steeg, S., Kapur, N., 2012. Premature death after self-harm: a multicentre cohort study. *Lancet* 380, 1568-1574.
- Bolton, J.M., Gunnell, D., Turecki, G., 2015. Suicide risk assessment and intervention in people with mental illness. *Bmj-British Medical Journal* 351.
- Carroll, R., Metcalfe, C., Gunnell, D., 2014. Hospital Presenting Self-Harm and Risk of Fatal and Non-Fatal Repetition: Systematic Review and Meta-Analysis. *Plos One* 9.
- Carter, G., Milner, A., McGill, K., Pirkis, J., Kapur, N., Spittal, M.J., 2017. Predicting suicidal behaviours using clinical instruments: systematic review and meta-analysis of positive predictive values for risk scales. *British Journal of Psychiatry* 210, 387-+.
- Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S., Devlin, N., Smith, P.C., Sculpher, M., 2015. Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technology Assessment* 19, 1-+.
- Cooper, J., Steeg, S., Gunnell, D., Webb, R., Hawton, K., Bennewith, O., House, A., Kapur, N., 2015. Variations in the hospital management of self-harm and patient outcome: A multi-site observational study in England. *Journal of Affective Disorders* 174, 101-105.
- Department of Health, 2015a. NHS reference costs 2014 to 2015. National schedule of reference costs: spell costs., 18/11/15 ed.
- Department of Health, 2015b. NHS reference costs 2014 to 2015. National schedule of reference costs: the main schedule.
- Department of Health, 2015c. Reference costs 2014 to 2015. National schedule of reference costs: spell costs.
- Devlin, N.J., Shah, K.K., Feng, Y., Mulhern, B., van Hout, B., 2017. Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health economics*.
- Erlangsen, A., Lind, B.D., Stuart, E.A., Qin, P., Stenager, E., Larsen, K.J., August G Wang, A.G., Hvid, M., Nielsen, A.C., Pedersen, C.M., Winsløv, J., Langhoff, C., Mühlmann, C., Nordentoft, M., 2014. Short-term and long-term effects of psychosocial therapy for people after deliberate self-harm: a register-based, nationwide multicentre study using propensity score matching. *Lancet Psychiatry* 2, 49-58.
- Haw, C., Bergen, H., Casey, D., Hawton, K., 2007. 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 and Life-Threatening Behavior* 37, 379-396.
- Haw, C., Hawton, K., 2008. Life problems and deliberate self-harm: Associations with gender, age, suicidal intent and psychiatric and personality disorder. *Journal of Affective Disorders* 109, 139-148.
- Haw, C., Hawton, K., Houston, K., Townsend, E., 2001. Psychiatric and personality disorders in deliberate self-harm patients. *British Journal of Psychiatry* 178, 48-54.

Haw, C., Hawton, K., Sutton, L., Sinclair, J., Deeks, J., 2005. Schizophrenia and deliberate self-harm: A systematic review of risk factors. *Suicide and Life-Threatening Behavior* 35, 50-62.

Hawton, K., Bergen, H., Casey, D., Simkin, S., Palmer, B., Cooper, J., Kapur, N., Horrocks, J., House, A., Lilley, R., Noble, R., Owens, D., 2007. Self-harm in England: a tale of three cities - Multicentre study of self-harm. *Social Psychiatry and Psychiatric Epidemiology* 42, 513-521.

Hawton, K., Bergen, H., Geulayov, G., Waters, K., Ness, J., Cooper, J., Kapur, N., 2016a. Impact of the recent recession on self-harm: Longitudinal ecological and patient-level investigation from the Multicentre Study of Self-harm in England. *Journal of Affective Disorders* 191, 132-138.

Hawton, K., Witt, K.G., Salisbury, T.L.T., Arensman, E., Gunnell, D., Hazell, P., Townsend, E., van Heeringen, K., 2016b. Psychosocial interventions for self-harm in adults. *Cochrane Database of Systematic Reviews*.

Hume, M., Platt, S., 2007. Appropriate interventions for the prevention and management of self-harm: a qualitative exploration of service-users' views. *Bmc Public Health* 7.

Kapur, N., Cooper, J., King-Hele, S., Webb, R., Lawlor, M., Rodway, C., Appleby, L., 2006. The repetition of suicidal behavior: A Multicenter cohort study. *Journal of Clinical Psychiatry* 67, 1599-1609.

Kapur, N., Steeg, S., Turnbull, P., Webb, R., Bergen, H., Hawton, K., Geulayov, G., Townsend, E., Ness, J., Waters, K., Cooper, J., 2015. Hospital management of suicidal behaviour and subsequent mortality: a prospective cohort study. *Lancet Psychiatry* 2, 809-816.

Kapur, N., Steeg, S., Webb, R., Haigh, M., Bergen, H., Hawton, K., Ness, J., Waters, K., Cooper, J., 2013. Does Clinical Management Improve Outcomes following Self-Harm? Results from the Multicentre Study of Self-Harm in England. *Plos One* 8.

Mars, B., Heron, J., Crane, C., Hawton, K., Lewis, G., Macleod, J., Tilling, K., Gunnell, D., 2014. Clinical and social outcomes of adolescent self harm: population based birth cohort study. *Bmj-British Medical Journal* 349, G5954-G5954.

Mathers, C.D., Vos, E.T., Stevenson, C.E., Begg, S.J., 2001. The burden of disease and injury in Australia. *Bulletin of the World Health Organization* 79, 1076-1084.

McDaid, D., Tsiachristas, A., Hawton, K., 2017. Understanding the true economic impact of self-harming behaviour Reply. *Lancet Psychiatry* 4, 901-901.

Miller, I.W., Camargo, C.A., Jr., Arias, S.A., Sullivan, A.F., Allen, M.H., Goldstein, A.B., Manton, A.P., Espinola, J.A., Jones, R., Hasegawa, K., Boudreaux, E.D., Investigators, E.-S., 2017. Suicide Prevention in an Emergency Department Population The ED-SAFE Study. *Jama Psychiatry* 74, 563-570.

Ness, J., Hawton, K., Bergen, H., Cooper, J., Steeg, S., Kapur, N., Clarke, M., Waters, K., 2015. Alcohol use and misuse, self-harm and subsequent mortality: an epidemiological and longitudinal study from the multicentre study of self-harm in England. *Emergency Medicine Journal* 32, 793-799.

NICE, 2011. The long term care and treatment of self-harm. Clinical Guideline 133. National Institute of Health and Care Excellence.

NICE, 2013. Guide to the methods of technology appraisal.

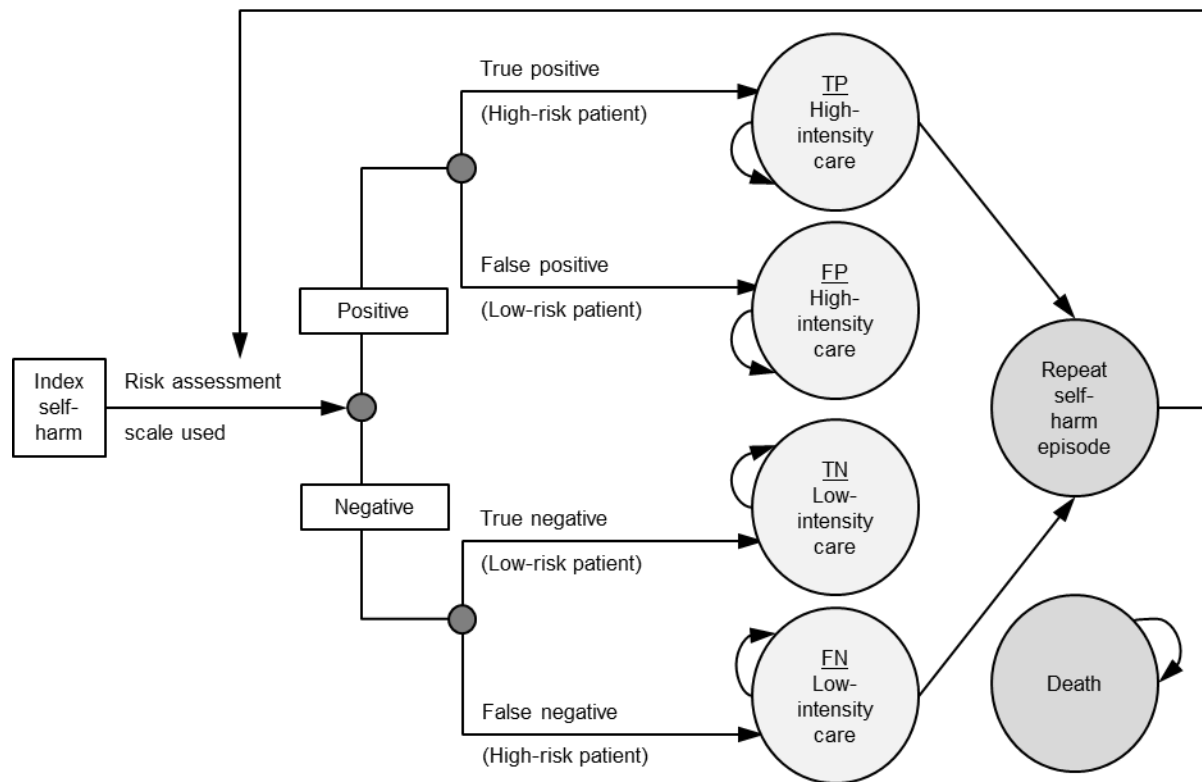
Quinlivan, L., Cooper, J., Davies, L., Hawton, K., Gunnell, D., Kapur, N., 2016. Which are the most useful scales for predicting repeat self-harm? A systematic review evaluating risk scales using measures of diagnostic accuracy. *BMJ open* 6, e009297-e009297.

Quinlivan, L., Cooper, J., Meehan, D., Longson, D., Potokar, J., Hulme, T., Marsden, J., Brand, F., Lange, K., Riseborough, E., Page, L., Metcalfe, C., Davies, L., O'Connor, R., Hawton, K., Gunnell, D., Kapur, N., 2017. Predictive accuracy of risk scales following self-harm: multicentre, prospective cohort study. *British Journal of Psychiatry* 210, 429-+.

Quinlivan, L., Cooper, J., Steeg, S., Davies, L., Hawton, K., Gunnell, D., Kapur, N., 2014a. Scales for predicting risk following self-harm: an observational study in 32 hospitals in England. *BMJ open* 4, e004732-e004732.

- Quinlivan, L., Cooper, J., Steeg, S., Davies, L., Hawton, K., Gunnell, D., Kapur, N., 2014b. Scales for predicting risk following self-harm: an observational study in 32 hospitals in England. *Bmj Open* 4.
- Royal College of Psychiatrists, 2010. Self-harm, suicide and risk: helping people who self-harm
- Final report of a working group. Royal College of Psychiatrists.
- Schwarz, G., 1978. ESTIMATING DIMENSION OF A MODEL. *Annals of Statistics* 6, 461-464.
- Sinclair, J.M.A., Gray, A., Rivero-Arias, O., Saunders, K.E.A., Hawton, K., 2011. Healthcare and social services resource use and costs of self-harm patients. *Social Psychiatry and Psychiatric Epidemiology* 46, 263-271.
- Sullivan, P.W., Slejko, J.F., Sculpher, M.J., Ghushchyan, V., 2011. Catalogue of EQ-5D scores for the United Kingdom. *Med Decis Making* 31, 800-804.
- Townsend, E., Ness, J., Waters, K., Kapur, N., Turnbull, P., Cooper, J., Bergen, H., Hawton, K., 2016. Self-harm and life problems: findings from the Multicentre Study of Self-harm in England. *Social Psychiatry and Psychiatric Epidemiology* 51, 183-192.
- Troquete, N.A.C., Van den Brink, R.H.S., Beintema, H., Mulder, T., van Os, T.W.D.P., Schoevers, R.A., Wiersma, D., 2013. Risk assessment and shared care planning in out-patient forensic psychiatry: cluster randomised controlled trial. *British Journal of Psychiatry* 202, 365-371.
- Tsiachristas, A., McDaid, D., Casey, D., Brand, F., Leal, J., Park, A.L., Geulayov, G., Hawton, K., 2017. General hospital costs in England of medical and psychiatric care for patients who self-harm: a retrospective analysis. *Lancet Psychiatry* 4, 759-767.
- van Spijker, B.A., van Straten, A., Kerkhof, A.J., Hoeymans, N., Smit, F., 2011. Disability weights for suicidal thoughts and non-fatal suicide attempts. *J Affect Disord* 134, 341-347.
- World Health Organization, 2014. Preventing suicide: A global imperative.

Figure 1: Cost-effectiveness model structure



The decision tree is on the left and occurs at the start of the model (index episode) to distribute patients into Markov health states on the right. High-risk patients may repeat self-harm over time, moving to the 'Repeat self-harm episode' state for 1 week. It is assumed that another risk assessment is conducted in the ED, which determines the health state into which the patient re-enters the Markov model. Patients can die at any point in the model and move to the 'Death' state.

Key: FN, false negative risk assessment (patient who repeats self-harm is incorrectly determined to be low-risk); FP, false positive risk assessment (patient who does not repeat self-harm is incorrectly determined to be high-risk); TN, true negative risk assessment (patient who does not repeat self-harm is correctly determined to be low-risk); TP, true positive risk assessment (patient who repeats self-harm is correctly determined to be high-risk).

Figure 2: Markov diagram showing the Clinician Global Rating (based on predictive ability performance in Quinlivan et al., 2017)

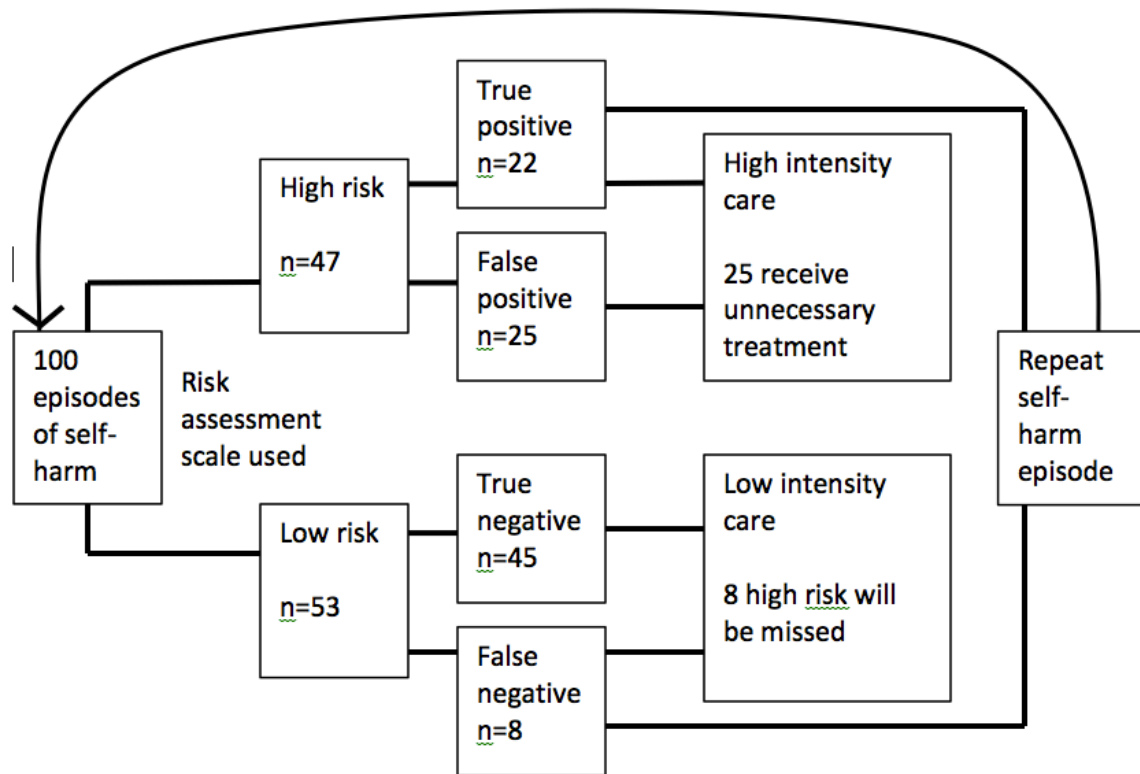
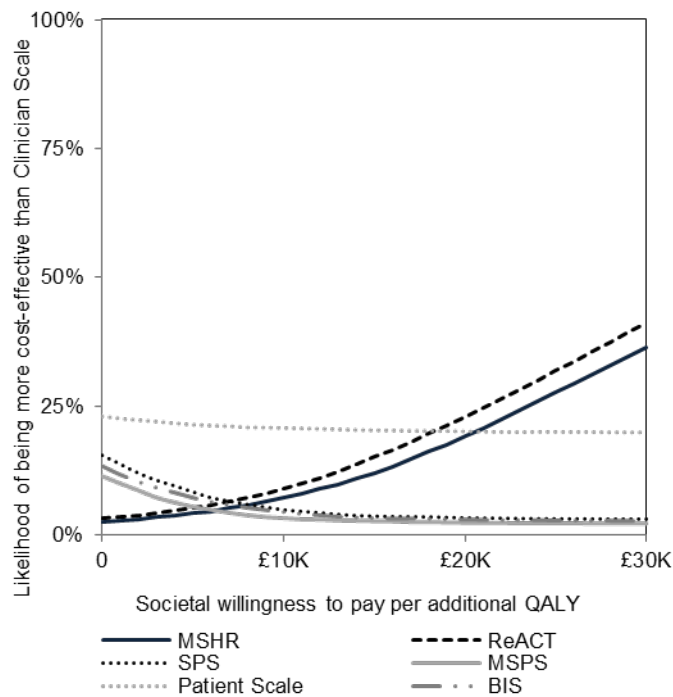


Figure 3: Cost-effectiveness acceptability curves for the risk scales



Note: 10,000 model runs were performed in order to produce smooth curves. 5,000 runs were used to inform the mean results so as to avoid artificially narrow confidence intervals caused by the number of observations (model runs).

Abbreviations: MSHR: Manchester Self-Harm Rule; ReACT: ReACT Self-Harm Rule; SPS: SAD PERSONS Scale; MSPS: Modified SAD PERSONS SCALE; Patient rating scale; BIS: Barratt Impulsivity Scale