

# International Journal of Offender Therapy and Comparative Criminology

<http://ijo.sagepub.com/>

---

## **Screening Tools Assessing Risk of Suicide and Self-Harm in Adult Offenders: A Systematic Review**

Amanda E. Perry, Rania Marandos, Simon Coulton and Mathew Johnson  
*Int J Offender Ther Comp Criminol* 2010 54: 803 originally published online 11 March  
2010

DOI: 10.1177/0306624X09359757

The online version of this article can be found at:  
<http://ijo.sagepub.com/content/54/5/803>

---

Published by:



<http://www.sagepublications.com>

**Additional services and information for *International Journal of Offender Therapy and Comparative  
Criminology* can be found at:**

**Email Alerts:** <http://ijo.sagepub.com/cgi/alerts>

**Subscriptions:** <http://ijo.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://ijo.sagepub.com/content/54/5/803.refs.html>

>> [Version of Record](#) - Sep 17, 2010

[OnlineFirst Version of Record](#) - Mar 11, 2010

[What is This?](#)

# Screening Tools Assessing Risk of Suicide and Self- Harm in Adult Offenders: A Systematic Review

Amanda E. Perry

Rania Marandos

*University of York, Heslington, United Kingdom*

Simon Coulton

*University of Kent, Canterbury, United Kingdom*

Mathew Johnson

*University of York, Heslington, United Kingdom*

This systematic review assessed the validity of screening instruments to identify the risk of suicide and self-harm behaviour in offenders. A search of 11 electronic databases and grey literature resulted in the inclusion of five studies. The five studies revealed four screening instruments, including the Suicide Checklist, the Suicide Probability Scale, Suicide Concerns for Offenders in Prison Environment (SCOPE), and the Suicide Potential Scale. Two instruments, SCOPE and Suicide Potential Scale, shared promising levels of sensitivity and specificity. The reporting of information was generally varied across items on the Standards for the Reporting of Diagnostic accuracy (STARD). Research is needed to assess the predictive validity of tools for offender populations in the identification of those at risk, particularly those in probation and community settings.

**Keywords:** *suicide; self-harm; systematic review; offenders*

Suicide and self-harm (SH) behaviour represents a major health problem within young adults under the age of 35 years. In 2005, the U.S. overall crude rate of suicide across all ages was 10.7 per 100,000 of the population. In this same year, almost double the number of males committed suicide between the ages of 20 and 35, suggesting that this age group is of particularly high risk (see <http://www.cdc.gov>).

**Authors' Note:** Address correspondence to Amanda E. Perry, Centre for Criminal Justice Economics and Psychology, University of York, Heslington, York, Y010 5DD, United Kingdom; e-mail: [aep4@york.ac.uk](mailto:aep4@york.ac.uk).

In the prison environment, this problem is exacerbated, with rates between 4 and 6 times higher than that of the general population and up to 8 times higher in newly released prisoners (Frühwald et al., 2002; Pratt, Piper, Appleby, Roger, & Shaw, 2006). In the United States, further distinctions between jails and prisons also acknowledge that rates of suicides in U.S. jails are greater than that in prisons mainly because of the nature of the population. With so many prisoners at risk, implications for the Criminal Justice System (CJS) focus on the clinical utility of how prisoners are identified. In such circumstances, screening needs to be quick, reliable, and accurate, producing as few false-positive results as possible (Bureau of Justice Statistics, 2005).

Existing prison screening procedures have been criticized as neither sensitive nor specific for detecting mental health problems and risk of suicide and SH behaviour (Mitichison, Rix, Renvoize, & Schweiger, 1994). Such criticisms highlight the need for a systematic programme of health screening to assess prisoners for mental health disorders and risk of suicide and SH behaviour (Birmingham, Gray, Mason, & Grubin, 2000). In addition, U.K. government targets aim to reduce suicide and SH behaviour by at least 20% by 2010 (Department of Health, 1998). Although these targets have been established, there is little evidence to suggest which methods of screening prevention would be effective in producing such a reduction (Guo, Scott, & Bowker, 2003).

## **Suicide and SH Risk Factors**

Suicide and SH behaviour is often linked to the presence of mental disorders, previous suicide and SH behaviour, and other psychological correlates such as drug and alcohol misuse and experiences of major life events (Arensault-Lapierre, Kim, & Turecki, 2004; Blaauw, Kerhof, & Hayes, 2005). Gender comparisons suggest that women are more likely to display SH behaviour whereas men are more likely to attempt suicide (Shaw, Baker, Hunt, Moloney, & Appleby, 2004; Skegg, 2005). Such problems tend to be highly prevalent within this already vulnerable prisoner population, making accurate identification of those most at risk somewhat problematic (Jenkins et al., 2005).

Practitioners in their assessment of those at risk should therefore consider factors that heighten and reduce potential risk. Research evidence tells us that entering prison for the first time, or being on remand, is known to put individuals at particularly high risk (Lader, Singleton, & Meltzer, 1998; Shaw et al., 2004). In addition to these known high-risk periods, the level of risk should be monitored continuously with the provision of regular after-care following an incident (Shaw et al., 2004). Protective risk factors aimed at reducing potential risk show that the presence of a child (particularly for women) is evidently important, and the effect of ethnicity (i.e., being Black) has been found to be inversely associated with suicide (Fazel,

Cartwright, Norman-Nott, & Hawton, 2008). Other reviews focusing on suicide assessment and screening tools with demographic, criminal, and psychiatric factors concluded that it might be feasible to identify prisoners at risk of suicide on the basis of a number of identified characteristics (Blaauw et al., 2005).

## Screening for Suicide and SH Behaviour

Screening for suicide or SH behaviour involves the classification of individuals using a classic two-by-two table identifying those who truly are at risk,  $a/(a + c)$  (i.e., the sensitivity of the instrument), and those who truly are not at risk,  $d/(b + d)$  (i.e., the specificity of the instrument).<sup>1</sup> The evaluation of any screening instrument therefore tends to involve a trade-off between the sensitivity and specificity of the instrument by manipulating the threshold or cut-off score that is used to identify a case.

This trade-off is used to maximize the likelihood that those who score positive have a high probability of being at risk, and those who score negative have a low probability of being at risk. Besides sensitivity and specificity, we are also interested in the positive predictive value (PVP;  $a/(a + b)$ ) and the negative predictive value (PVN;  $d/(c + d)$ ) of the screening tools. The PVP is the proportion of individuals who are identified by the screening tool as being at risk of suicide or SH behaviour who then actually go on to display such behaviour. Conversely, the PVN is the proportion of people identified by the screening tool as not at risk who do not go on to display the behaviour.

Adverse effects from screening occur from the misclassification of individuals. For example, the misclassification of a false-positive result could lead to unnecessary investigation and treatment. False-negative predictions could lead to a failure of the CJS to provide support where it is needed. Because of the fatality implications of not identifying someone at risk of suicide correctly, the optimum trade-off occurs when the number of false negatives is very small (i.e., the screening does not miss anyone who may be at risk). In contrast, assessment of suicide and SH behaviour would usually be conducted once someone had been identified as being at risk. Such assessment would be used to obtain an in-depth understanding of the individual, including his or her risks and previous behaviours.

In addition to these considerations, a number of other problems are highlighted in the literature. These problems include the transferability of existing scales (developed originally in psychiatric populations), the prevalence of suicide within the population, and the lack of an apparent gold standard test (Correia, 2000; Mills & Kroner, 2000; Perry & Olason, 2009).

The transferability of existing screening and assessment instruments are well recognized for screening suicide and SH risk in general and psychiatric populations (e.g., Beck Depression Inventory [BDI] and the Beck Hopelessness Scale [BHS]). Simply importing scales developed in one population (psychiatric outpatient) to a different

population (i.e., prisoner population) is problematic because of the unique environment within which prisoners are accommodated (Correia, 2000; Mills & Kroner, 2000; Perry, 2005). For example, two items on the BDI-II investigate feelings of guilt (Item 5, "I feel particularly guilty") and punishment (Item 6, "I feel I am being punished"). Elevated scores on these two items in a prison population are perhaps not solely a reflection of inherently high levels of depression among prisoners but also a result of the unique circumstances associated with committing a crime (i.e., feelings of guilt) and being in a prison environment (i.e., feelings of punishment).

Prevalence is another practical problem of screening for suicide and SH behaviour in the prisoner population. Although the prevalence of suicide is greater than that in the general population, it still remains a relatively rare event. Such rare events make prospective prediction problematic as it requires national longitudinal studies to examine the predictive validity of an instrument. In such circumstances, proxy measures such as SH behaviour, or number of previous suicide attempts, are used to assess the accuracy of such measures.

The final problem highlights the lack of gold standard tests. A gold standard test or criterion standard test is one that is recognized as a diagnostic test or benchmark. Traditional psychometric methods compare the use of an index test (i.e., the tool under investigation) to a gold standard test. A hypothetical gold standard test is one that has 100% specificity and sensitivity. In practice, there are no ideal gold standard tests and gold standard tests can change over time as new technologies provide more advanced methods of assessment. Because gold standard tests can be incorrect (i.e., they can misclassify), the results of any test evaluation should be interpreted in the context of the population and the test setting. When a gold standard test is absent, a reference test or imperfect gold standard test can be used instead. Although well-established methods are used for assessing the accuracy of screening tests when gold standard tests are available, such methods are limited when gold standards are absent (Alonzo & Pepe, 1998).

Studies investigating screening efficacy when gold standard tests are lacking tend to use populations of people who are known to be at risk, by comparing them with people with no previous risk. However, measuring the true risk status in this fashion does not provide certainty because risk measured using historical information may be fraught with inaccuracies and may not be highly predictive of future behaviour (Arboleda-Florez & Holley, 1989; Liebling, 1992). In health care, the absence of such gold standard tests has been addressed with the development of gold standard-discrepant resolution analysis (e.g., Cullen, Long, & Lorincz, 1997) and use of composite reference standard tests and latent class analysis (Faraone & Tsuang, 1994).

Suicide and SH behaviour represent a major health problem in the prisoner population. Government targets aimed at reducing the incidence of suicide are established with little known evidence about how to identify those most at risk. A number of researchers have investigated the relationship between suicide and SH

behaviour and have identified a list of known risk factors. The results of such studies show differences according to gender, age, and sociodemographic factors. However, screening for such risk factors is problematic for a number of reasons. These include the lack of validated instruments, the low prevalence of suicide, and the lack of apparent gold standard tests. Screening instruments for measurement of suicide and SH behaviour exist but a systematic review of this literature has not been conducted. This article reports on the findings from a systematic review evaluating the validity of the screening accuracy of assessment tools used to identify suicide and SH risk in adult offenders.

## Method

The systematic review aimed to evaluate studies assessing the validity of screening and assessment tools to identify risk of suicide and SH behaviour in adult offenders. Studies included in the review were identified using four key criteria and had to include (a) an assessment of risk for suicide or SH behaviour using a screening tool (any screening tool suitable for use within the criminal justice system was eligible), (b) a study sample with a mean age of 35 years or less (this age group was used primarily because adult offenders within this age bracket are known to be at particularly high risk of suicide and SH behaviour within the prison population; Samaritans, 1998), (c) a sample of offenders under the care of the criminal justice system (these included offenders under the jurisdiction of the court system, residential and secure establishments, or in the community on parole or probation), and (d) a test of the reliability or validity of a screening tool.

### Search Strategy for Identification of Studies

A search strategy was devised to identify published and unpublished literature. Searches were not restricted to any single language or nationality. A preliminary investigation of the literature identified studies from 1980 onwards as this was found to be when most of the literature had been subsequently published. The published literature searches were conducted on 11 databases between January 1980 and June 2001 and between January 1980 and November 2004.<sup>2</sup> The search for unpublished literature was conducted separately but used the same inclusion and exclusion criteria as devised for the published literature. Six key sources of information were identified for searching.<sup>3</sup> Keyword search terms were divided into five different areas, including details about screening, the population, the subject of suicide and SH, and the setting.<sup>4</sup>

### Assessment of Information Reported in the Studies

The studies were assessed using the STAndards for the Reporting of Diagnostic accuracy studies statement (STARD; Bossuyt et al., 2003). Using the STARD, each

study was rated a three-point scale. Studies rated as “yes” reported all information in the descriptive methodological item and were given a score of 1. Studies rated as “partial” reported some but not all aspects described in the descriptive methodological item and were given a score of 0.5. Studies rated as “no” reported none of the aspects described in the descriptive methodological item and were given a score of 0. For each study, the reporting of information was assessed using two independent reviewers. Where inconsistencies were reported these were discussed until agreement was reached.

### **Data Management and Extraction**

Studies were screened initially by the key author and subsequently by two independent reviewers. Data were independently extracted from studies using a specifically designed data extraction sheet (see Perry & Marandos, in press, for full details). Missing data were obtained from the authors via letters, facsimiles, telephone conversations, and e-mail. Any discrepancies were subsequently resolved by referring to the source material. The information was stored electronically in a Microsoft Access Database and Endnote was used as a bibliographic management system for storing, locating, and tracking references.

### **Data Synthesis**

The data synthesis was conducted in two stages. First, narrative tables describing the characteristics of the study participants and instrument details were constructed. Table 1 includes information on the details of the study (i.e., author and year of publication), sample size and study setting, age of the participants, the screening or index test used in the study, and the comparison standard or criterion reference test. Second, for each scale, we constructed a series of two-by-two tables. From these we calculated the sensitivity, specificity, PVP, and PVN together with the misclassification rate of each study (see Table 3). A meta-analysis was not completed because of the heterogeneous nature of the study designs (i.e., no two studies examined the same screening and assessment tool with a similar population of offenders).

## **Results**

The search methods generated 404 published citations and 75 pieces of grey literature. All unpublished literature was primarily excluded because none assessed the validity of a screening tool. Following this initial pre-screening, a total of 25 potential published citations were identified for further examination. A secondary pre-screening was independently conducted by two reviewers. Inter-rater reliability was conducted on the first six studies and an average 82.5% agreement was obtained. The range of agreement was between 70% and 95%.

**Table 1**  
**Description of All Primary Studies**

First author (year), country	Sample size (N)	Setting	% Male	M age (SD)	Screening test	Comparison standard	Study design
Arboleda-Florez (1988, 1989), Canada	141	Remand centre	89	Not reported	Suicide Checklist	Actively suicidal ( $n = 14$ ) vs. history of suicide ( $n = 82$ ) vs. no history of suicide ( $n = 45$ )	Screening on a consecutive series of admissions after the target event
Daigle (1999), Canada	88	2 provincial and 1 penitentiary prison	0	Overall group, 33 years (9.0)	Suicide Probability Scale	Known history of suicide attempt ( $n = 47$ ) vs. no known history ( $n = 41$ )	Screening on a volunteer sample using concurrent self-report data about the targeted event
Earthrowl (2002), New Zealand	150	1 female offender prison	0	Overall group, 29.6 years (SD not reported)	Suicide Checklist	Known history vs. ( $n = 34$ ) no known history ( $n = 90$ )	Screening on a series of admissions after the targeted event
Perry (2005), United Kingdom	1029	6 young offender institutions	60.1	Known history, 24.8 years (8.99) No known history, 23.4 years (9.56)	Suicide and Self-Harm Concerns for Offenders in Prison Environment (SCOPE)	Known history of vulnerability to risk of suicide or SH behaviour ( $n = 241$ ) vs. no known history ( $n = 747$ )	Screening on a purposive sample using concurrent data
Wichmann (2000), Canada	152	Adult offender establishment	100	Suicide attempt group, 23.88 years (SD not reported) No known history, 23.91 years (SD not reported)	Suicide Potential Scale	Known history of suicide attempt ( $n = 76$ ) vs. no known history ( $n = 76$ )	Screening on a sample of offenders prior to the targeted event



Of the 25 studies, 19 were subsequently excluded for a number of different reasons. Four studies (Battle, Battle, & Tolley, 1993; Falloon, 2000; Grisso, Barnum, Fletcher, Cauffman, & Peuschold, 2001; Stein & Yeager, 1993) were excluded because they did not include a criterion group of offenders. Another four studies (Boothby & Durham, 1999; Defrancesco, Armstrong, & Russolillo, 1996; Huckaby, Kohler, Garner, & Steiner, 1998; Inch, Rowlands, & Soliman, 1995) were excluded because they did not focus on the prediction or postdiction of suicide and SH risk. One study (Blaauw, Kerhof, Winkel, & Sheridan, 2001) contained no validity data, and two further studies (Blaauw et al., 2005; Rohde, Seeley, & Mace, 1997) did not focus on the psychometric properties of an instrument but instead presented a composite test battery of measures. The final eight studies did not provide enough information to calculate the sensitivity and specificity of the instruments (Archer, 1989; Cole, 1989; Esposito & Clum, 1999; Kempton & Forehand, 1992; Ivanoff & Jang, 1991, 1994; Ivanoff, Smyth, Grochowski, & Jang, 1992; Phillips & Dahlstrom, 1997).

### Description of the Studies

The six studies included in the review were published between 1989 and 2005 and were conducted in Canada ( $n = 4$ ), New Zealand ( $n = 1$ ), and the United Kingdom ( $n = 1$ ). Of the six studies, two reported on the successive development of the Suicide Checklist (SCL; Arboleda-Florez & Holley, 1988, 1989) and one further study reported on the use of this same checklist (Earthrowl & McCully, 2002). For the purpose of data extraction these three articles were subsequently treated as two studies (i.e., the Arboleda-Florez studies were combined as one). This left five separate studies for extraction (see Appendix). The five studies represented four screening measures: the Suicide Probability Scale, Suicide Potential Scale, SCL, and Suicide Concerns for Offenders in Prison Environment (SCOPE).

Table 1 shows the variation in demographic characteristics across study participants. The samples were drawn from a variety of settings. These included male and female offenders in a remand centre in Canada (Arboleda-Florez & Holley, 1989), six young offender institutions (one female and five male) in the United Kingdom (Perry, 2005), female offenders in two provincial and one penitentiary prison in Canada (Daigle, Alarie, & Lefebvre, 1999), female offenders in a prison in New Zealand (Earthrowl & McCully, 2002), and male offenders in a young adult institution in Canada (Wichmann, Serin, & Motiuk, 2000). The ethnic composition across four of the five studies ranged from 68% to 84.4% White. One study contained a majority Maori sample (Earthrowl & McCully, 2002), whereas another did not report on the ethnic origin of the sample (Daigle et al., 1999). The mean age range of participants across the five studies was between 23 and 33 years of age. One study did not report a mean age and instead referred to them as "young offenders." This study was therefore subsequently accepted in the review (Arboleda-Florez & Holley, 1989). The size of the samples varied between 88 and 1,029.

## Screening Measures

*SCL.* Two studies reported on the development and validation of the SCL in men and women (Arboleda-Florez & Holley, 1988, 1989), and one study reported on the use of the SCL in a population of female offenders (Earthrowl & McCully, 2002). Together, the information presented in these three studies provides information about the psychometric properties of the instrument.

The SCL consists of 11 clinical items of symptoms relating to current depression and suicidal ideation and 6 historical items relating to previous suicidal or problem behaviours and any relevant psychiatric complaints. The SCL was developed to identify acutely distressed inmates requiring immediate intervention, placement, observation, and resource allocation. The checklist was administered by nursing staff but can also be administered by prison staff with only brief training. The authors note that the tool is not a substitute for more detailed diagnostic assessment. The 11 clinical items are scored 0 (*not in evidence*), 1 (*somewhat in evidence*), and 2 (*very much in evidence*), creating a possible maximum score of 22. The historical items are scored 0 (*absent*) or 1 (*present*), creating a maximum score of 6. Most interviews are completed within 8 to 10 min.

Two different cut-off scores were presented in the articles. The Arboleda-Florez and Holley study (1989) found that symptom scores of 4 or more indicated greatest risk but was somewhat over-inclusive. The false-positive rate is not listed by the authors, so it is difficult to assess how over-inclusive this threshold would be. The Arboleda-Florez and Holley study showed the SCL to be a good but over-inclusive measure of risk (i.e., it identifies large numbers of false positives). The authors concluded that greater accuracy could be achieved if the symptom score was associated across different groups of offenders on the basis of certain criminological variables. These included whether individuals were divorced or separated, prisoners who were sentenced, and those with recorded weapons offences.

The Earthrowl and McCully study raised the cut-off score from 4 to 6. This resulted in increasing the number of missed cases from 10 to 17. In this study, this produced a false-negative result of 41%. This study found that the SCL moderately discriminated between the acutely distressed prisoners in need of immediate intervention (requiring placement within the prison, observation, and resource allocation) in comparison with those who may require long-term ongoing care management during a period in custody. Using a cut-off score of 6 or more on the symptoms scale, the tool had 70% sensitivity and 21% specificity. This means that little less than 29% of the sample will be falsely identified as being at risk (i.e., false positive), and less than 21% will be identified incorrectly as being not at risk when they are (i.e., false negative). The overall misclassification rate for these two combined is 25%.

*The Suicide Probability Scale.* One study reported on the validation of the Suicide Potential Scale with female offenders from two prisons in Quebec

(Daigle et al., 1999). This scale was developed in the United States as a self-report paper-and-pencil screening instrument designed to be used with adults (Cull & Gill, 1988). It contains 36 items and assesses four areas: hopelessness, suicidal ideation, negative self-evaluation, and hostility. Administered in groups or individually, the scale takes between 5 and 10 min to complete. Respondents are instructed to circle whether each item on the scale describes them as *none or a little of the time*, *some of the time*, *good part of the time*, or *most or all of the time*. Interpretation of the scale is based on individual item-analysis scores on the four subscales and the total weighted score (and *T* score). The authors suggest that the Suicide Probability Scale should not be used as the sole instrument for assessing suicidality and that it is meant to supplement clinical judgment.

Test-retest reliability ( $r = .92$ ) and internal consistency ( $r = .94$ ) of the Suicide Probability Scale was found to be high in the original development of the scale (Cull & Gill, 1988). Overall, the instrument had a sensitivity of 53% and a specificity of 78%. This means that slightly less than 50% of those identified will be falsely identified as being at risk (i.e., false positive) and less than 25% will be identified as not at risk when they are (i.e., false negative). The overall misclassification rate for this instrument is 35%.

*SCOPE*. One study reported on the SCOPE with male and female offenders in the United Kingdom (Perry, 2005). The SCOPE was developed as a self-report paper-and-pencil screening instrument designed to identify risk of suicide in young adult prisoners. It consists of 28 items across two factors referred to as protective social network and optimism. The protective social network factor relates to the presence of a social support network (e.g., family, peers, and prison staff), innate coping, and use of problem-solving strategies such as talking to someone when you have a problem. The optimism factor reports on symptoms of depression, hopelessness, SH behaviour, and suicide intent. The protective social network factor contains 12 items, giving a maximum score of 72. The optimism factor contains 15 items, giving a maximum score of 90. This self-report tool is administered on an individual basis and can be quickly administered by general prison staff, taking only a few minutes to complete. Respondents are asked to rate their responses on a 6-point Likert-type scale from *strongly agree* (rated as 0) to *strongly disagree* (rated as 5). The tool primarily identifies those with the greatest predisposition toward suicide and SH risk rather than suicide per se. Scores greater than 38 (on Factor 1) and scores greater than 30 (on Factor 2) are indicative of individuals displaying predispositions toward risk of suicide and SH behaviour.

The internal consistency of the scale on each factor was good (.72 and .85), and SCOPE discriminated between those with a history of suicide or SH behaviour and those with no known history ( $p < .01$ ). Those at risk on average scored significantly higher, 55 ( $SD = 14.85$ ), than those not at risk, 37 ( $SD = 10.93$ ), with 95% confidence intervals (-19.9 to -16.2). Overall, the SCOPE reported 81% sensitivity and 71% specificity with a cut-off score of  $\geq 78$ .

*Suicide Potential Scale.* One study discussed the evaluation of the Suicide Potential Scale with a group of male offenders in Canada (Wichmann et al., 2000). This scale was developed to standardize practice among staff members dealing with offenders currently expressing suicidal ideation and determining immediate risk of engaging in suicidal behaviour. The scale contains nine items and assesses previous suicide history, recent psychological or psychiatric intervention, recent loss of relative or spouse, current major problems, influence of alcohol and drugs, signs of depression, expressed suicidal ideation, and presence of a suicide plan. No details about the administration of this instrument are provided. Each of the nine items are scored as 1 (*present*) or 0 (*absent*). This provides a maximum score of 9. The higher the mean score, the more likely an individual is to be at risk. The mean number of indicators endorsed by people with a previous history of suicide attempts was 1.61 in comparison to 0.33 in a group of people without any previous history of suicide attempt.

The internal consistency of the scale was good (.77) and also demonstrates the ability to discriminate between attempters and non-attempters. The Suicide Potential Scale (Wichmann et al., 2000) showed that five variables correctly identified 92% of cases. The model produced a false-positive rate of 14% and a false-negative rate of 20%.

*Assessment using the STARD.* Table 2 shows the findings of the STARD assessment. Of the 25 items, 10 items (40%) scored 1 or less on the STARD. These included 4 items scoring nothing (Items 1, 13, 19, and 24), 2 items scoring 0.5 (items 10 and 11), and 4 items scoring 1 (Items 16, 17, 20, and 22). Of the remaining 15 items, 7 items scored  $\geq 2$  (Items 2, 4, 12, 14, 18, 21, and 23) and 7 items scored  $\geq 4$  (Items 3, 5, 7, 8, 9, 15, and 25); the final item (Item 6) referred to the design of the study and was therefore not scored. Overall, the reporting of information in these studies was generally poor, with many studies not reporting much information on the STARD. Of particular concern is the lack of detail provided about the estimates of diagnostic accuracy. Ideally, such studies should enable the calculation of different estimates of risk prevalence and varying cut-off points. Such information would result in diagnostic odds ratios, enabling the area under the curve (AUC) statistics to be generated for receiver operating characteristic curves.

## Screening Efficacy

Table 3 shows the screening efficacy across all four instruments. The sensitivity (53% to 86%) and specificity (21% to 80%) values varied across the instruments. Two studies did not present enough information in the two-by-two tables to calculate the PVP and PVN (Arboleda-Florez & Holley, 1989; Wichmann et al., 2000). For the other three studies, the figures for the PVP varied from 74% to 94% and from 24% to 78% for the PVN. The overall misclassification rate ranged from 17% to 35%.

**Table 2**  
**Reporting of Information Assessment**

First author (year)	Title and introduction		Participants			Test methods			Statistical methods			Participant results			Test results			Estimates			Discussion				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		21	22	23	24
Arboleda- Florez (1988, 1989)	0	0.5	0.5	0	1	Retrospective	1	0.5	0.5	0.5	0.5	0	0	0	1	0	0	0	0	0	1	1	0.5	0	25
Daigle (1999)	0	0	0.5	1	0.5	Concurrent	0	0.5	1	0	0	0	N/A	0	1	0	0	0.5	0	0	0.5	0	0.5	N/A	1
Earthrowl (2002)	0	0.5	1	1	1	Retrospective	1	1	1	0	0	0	0	1	1	0.5	0	1	0	1	0	0	0	0	1
Perry (2005)	0	1	1	0.5	1	Concurrent	1	1	1	0	0	1	0	1	1	0	1	1	0	0	1	0	0.5	0	1
Wichmann (2000)	0	1	1	1	1	Postdiction/ Predictive	1	1	0.5	0	0	1	0	0	0	0.5	0	0	0	0	0	0	0	0	1

Note: 0 = not reported; 0.5 = partial information reported; 1 = full information reported.

**Table 3**  
**Sensitivity and Specificity Values**

First author (year)	Predicting	Screening tool (cut-off score)	Sensitivity (%)	Specificity (%)	PVP (%)	PVN (%)	Overall misclassification rate
Arboleda-Florez (1989)	Men with history of attempted suicide Men with current suicidal ideation	SCL (>4)	86	Not reported	Not reported	Not reported	Not reported
Earthrowl (2002)	Women with a history of previous suicide	SCL (>6)	70	21	74	24	25
Daigle (1999)	Women at risk of attempted suicide or SH behaviour	Suicide Probability Scale (moderate and high risk)	53	78	74	78	35
Perry (2005)	Men and women with vulnerability to risk	SCOPE (>78)	81	71	94	55	17
Wichmann (2000)	Male suicide attempters	Suicide Potential Scale (not stated)	86	80	Not reported	Not reported	Not reported

Note: PVP = predictive value, positive; PVN = predictive value, negative; SCL = Suicide Checklist; SH = self-harm; SCOPE = Suicide Concerns for Offenders in Prison Environment.

The sensitivity of the instrument refers to the identification of those truly at risk. Three instruments had high levels of sensitivity. These included the SCL, with a cut-off of  $>4$  (86%); the SCOPE, with a cut-off of  $>78$  (81%); and the Suicide Potential Scale (86%). These three instruments were used on different populations, identifying male offenders on entry to a remand centre in Canada (SCL), adult male and female offenders in adult prisons in the United Kingdom (SCOPE), and male offenders who had previously attempted suicide in a federal prison in Canada (Suicide Potential Scale).

Decisions about choosing which instrument to use are not simply based on the instrument's ability to identify true-positive cases correctly. For example, these three studies represent different screening populations with differing levels of risk (i.e., men on entry to a remand centre are likely to be at greater risk than women and men in a young offender prison).

Therefore equally important are the setting for screening, the prevalence of risk within the population, and correct identification of those who are truly not at risk. Four of the five studies reported on the specificity of the instruments (i.e., correctly identifying someone who is not at risk). These included the SCOPE (71%), the Suicide Probability Scale (78%), the Suicide Potential Scale (80%), and the SCL (21%). The combined evaluation of any screening instrument involves the trade-off between the sensitivity and specificity of the instrument and the threshold or cut-off score used to identify someone at risk. In our four instruments, the Suicide Probability Scale had good specificity (78%) but only moderate sensitivity (53%). The reverse is true of the SCL, where a cut-off score greater than 6 indicates good sensitivity (70%) but poor specificity (21%). Ultimately, the balance between the sensitivity and specificity of an instrument should be optimized by altering the threshold or cut-off score. This allows correct identification of the maximum number of individuals who are, and are not, at risk. Improvements in specificity are only obtained by raising the cut-off point. With this regard, the SCOPE (sensitivity 81% and specificity 71%) and the Suicide Potential Scale (sensitivity 86% and specificity 80%) fare better than the SCL (sensitivity 70% and specificity 21%) and the Suicide Probability Scale (sensitivity 53% and specificity 78%).

In taking the sensitivity and specificity of an instrument into consideration, the PVP and the PVN are also important. The PVP is the probability that a person is actually at risk of suicide and SH behaviour given that he or she tests positive. The predictive value of a screening test is determined not only by the validity of the test (i.e., sensitivity and specificity) but also by the characteristics of the population and the prevalence of risk within the population. For rare events such as a suicide, a major determinant is the prevalence of risk within the screened population. No matter how specific the test, if the population is at low risk of actually committing suicide, results that are positive will mostly be false positives. In our set of five studies, only one study (Arboleda-Florez & Holley, 1988, 1989) reported on the prevalence rates (28.2 per 1,000 for actively suicidal individuals and 165.3 per 1,000 with a known history of past suicide). For the other four studies, the percentage of individuals identified as being at risk across the

studies varied from 33% to 53%. Such high levels of risk pose screening problems of identifying those most likely to go on and present with suicidal behaviour.

In terms of the PVP, the SCL and the Suicide Probability Scale had the same PVP (74%), with the SCOPE faring best (94%). For the SCL and the Suicide Probability Scale, this means that 26% of all individuals who were referred for further assessment were not actually found to be at risk. In a population of 1,000 people, this would mean that 260 individuals would have been assessed and/or resources allocated inappropriately. The PVP can be increased by increasing the specificity of the test (by changing the criterion of positivity) or by increasing the prevalence of risk in the screened population. This can be accomplished by targeting the screening program to groups of individuals at differing levels of risk.

The reverse is true for the PVN. The PVN is the probability that a person is truly risk free given a negative screening test. The more sensitive a test, the less likely it is that an individual with a negative test will be at risk and thus the greater the PVN. For the SCL, only 24% of individuals were classified in this manner. In comparison, the SCOPE (55%) and the Suicide Probability Scale (78%) fared significantly better. A high PVN is to be expected in any screening program for an event that is rare because the vast majority of those screened will by definition be risk free. As a consequence, the PVP and the PVN are affected by the outcome measure used to identify the level of risk within the population. For example, using suicide as an outcome measure would result in a high PVN. For SH, the PVP and the PVN would be different as many more people under the care of the CJS harm themselves than those that commit suicide, and the PVN would be considerably lower because the majority of those screened will not be risk free.

## Discussion

The five studies identified in this review were conducted over an 11-year period. This resulted in the identification of four screening instruments developed across three different countries. None of the instruments were identified as assessment tools and all were used to identify the probability of someone being at risk (as opposed to being diagnosed). The conclusions on which this review is based focus on the identification of problems associated with screening for suicide and SH behaviour in an offender population. These problems include the impact of the study design and the importance of predictive validity. Second, the balance of sensitivity and specificity of an instrument, the use of these screening tools in different screening populations, and the reporting of information in the included studies.

Two of the five studies used retrospective methodology to consider the identification of individuals at risk. Limitations of retrospective designs are usually associated with the collection of data that occurred previously. Such data may be recorded for a different purpose; this can result in incomplete or possibly non-comparable information for all study participants. Perhaps more important is the issue that none of the five studies



assessed the predictive validity of future suicide and SH behaviour. Predictive validity studies are inherently difficult to conduct in transient and vulnerable populations. Such studies are also expensive to conduct, but without them we cannot address the main epidemiological question surrounding the validity of screening, that is, whether prisoners fare better as a consequence of screening in comparison to those not tested.

Other aspects of validity concern the sensitivity and specificity of the instruments. Two of the four instruments (SCOPE and Suicide Potential Scale) reported high levels of sensitivity and specificity, generating an acceptable optimum trade-off between the two. Ultimately, decisions about choosing which is most valid depend on the characteristics of the screening population, the setting (e.g., offenders on remand are at greater risk than those who are sentenced), the prevalence of risk within the population, and the proposed cut-off identified for "caseness." Each of our studies used different populations of offenders in different settings. Only one tool was used in different populations: the SCL, which was used with Maori female offenders in New Zealand and male and female offenders in a remand centre in Canada. Such findings for the SCL demonstrate some evidence toward its use in different populations. For the remaining three instruments (SCOPE, Suicide Potential Scale, and Suicide Probability Scale), the external validity or generalizability of these screening tools with other populations of offenders across different countries is as yet unknown. In other screening tools, cultural differences have been responsible for altering the cut-off scores used in the United Kingdom and United States (e.g., Cooke, 1995).

The balance between the sensitivity and the specificity of the instrument must also be considered in relation to the proposed outcome measure. In the identification of suicide, one might for example argue that the specificity of the test is more important than the sensitivity, as missing someone who is at risk of suicide is fatal. Increasing the specificity of a test would enable the exclusion of many who are not truly at risk of attempted suicide. However, the prevalence of suicide within the population is relatively rare, and prospective study designs would require long follow-up periods to assess the predictive validity of such instruments.

Using SH as an outcome, our sensitivity levels and specificity may look quite different as many more people are likely to be at risk. In such cases, the trade-off between sensitivity and specificity could be more equally balanced, as the rate of false positives and negatives could be taken into consideration with the amount of available resources. Given that all of our four instruments focused on previous attempted suicide or SH risk, the balance of sensitivity and specificity seems to be most important. With this regard, the SCOPE and Suicide Potential Scale resulted in acceptable levels of sensitivity and specificity. The SCOPE was used specifically with male and female offenders in prisons in the United Kingdom, and the Suicide Potential Scale was used with male offenders in a remand centre in Canada. Further validation of the Suicide Potential Scale is needed to assess its use with female offenders. This is especially relevant given the increasing numbers of incarcerated women and the high levels of SH behaviour within the female prisoner population (Shaw et al., 2004).

Despite these acceptable levels of sensitivity and specificity, the reporting of information in the studies was generally poor, with all studies having some elements of reporting on the STARD missing. This included information about the test execution, rationale for the cut-off scores, and the expertise and blinding of test administrators. In addition, descriptions about the statistical methods and the lack of diagnostic estimates lead to concerns about the reproducibility and external validity of these findings with other prisoner populations.

## Implications for Practitioners

There are a number of practical implications that arise from this research. First, an important consideration for practitioners is the clinical utility (i.e., the ease and practicality of administration). Of the screening instruments, all of the tools identified in this review could be administered within a short period of time. A number of different staff members used these tools, including nursing staff, prison officers, and staff on reception into a prison setting. Information that was not provided in the studies related to the requirements for staff training. This means that although many different staff members need to have the ability to screen for differing levels of risk of suicide and SH behaviour, we have little knowledge about what training would be needed to implement a screening program in the CJS.

Second, some tools were found to be more specific (identifying those who are truly not at risk) than others (e.g., Suicide Potential Scale and the Suicide Probability Scale). Levels of specificity in the CJS are an important concern given that resources need to be appropriately targeted. Linked to this second concern is the issue of false positives (i.e., identifying individuals at risk when they are not). On one hand, CJS screening tools need to minimize the number of people incorrectly identified to target the appropriate resources to those who most need the support and advice. However, this approach needs to be balanced by the number of people classified as a false negative.

There are limitations in how much we can ascertain from the results of these studies in a practical application. As such, it is not known how stable the sensitivity, specificity, and other measures of diagnostic accuracy are across offender groups and CJS settings. Nor do we know whether these tools are able to identify a change in mood or level or risk over time. Furthermore, none of the instruments were validated with offenders in the community who were under the care of the probation service, courts, or the police. Such validation is required to improve the identification of those at risk while in custodial care of the CJS. It should also be noted that none of these four instruments were referred to as assessment tools. All were validated as screening instruments to identify well individuals from those who are probably not. These instruments are therefore not meant to be diagnostic; rather, they seek to separate an at-risk group of individuals so that they can be referred for more detailed assessment and treatment. With this in mind, some over-inclusiveness (i.e., false-positive results) may therefore be tolerable.

## Implications for Research

There is a clear need for additional psychometric research on the validity of suicide and SH behaviour screening tools in offender populations. Some of the challenges outlined at the beginning of this discussion still remain problematic. These include the use of appropriate outcome measures, the lack of any gold standard test, and the identification of a group of individuals who exist within an already vulnerable population. As an outcome measure, suicide is not a practical or feasible measure given that 97 individuals in a prison population of 78,000 is a typical example of the number of suicides within the United Kingdom. Nevertheless, other measures for suicide, such as the number of SH incidents, may serve as a proxy indicator given what we know about the levels of increased risk of suicide in people who harm themselves. Studies are needed to evaluate screening tools in different settings in the CJS and should consider whether such tools are able to measure changes in level of risk and mood over time. Methodologically, the reporting of information in the studies was generally poor. Improvements in the reporting of information would be encouraged if STARD guidelines were stipulated by commissioners of research and journal editors. These requirements would aid the design and reporting of research studies and ultimately improve the quality of research in the criminal justice system. The challenge for research that still remains is whether prisoners fare better as a consequence of screening in comparison to those not tested. As yet, this question remains unanswered (Sackett & Haynes, 2002).

## Recommendations

A number of recommendations arise from this research. The main recommendations cover the concerns for establishing policies around the prevention of future suicide and SH behaviour in prisons. Without such evidence, such policies are difficult to develop given that it is not yet known how effective these tools are in predicting the future risk of suicide and SH behaviour and what can be done to help these individuals once they have been identified. Any successful screening program is determined by the ability to then provide an effective intervention for those at risk. Without this, some would argue that screening is a waste of resources. For practitioners, requirements for training in the use of such instruments should be developed as part of the protocol and implementation of any screening program (Hayes & Lever-Green, 2006; Shaw, Appleby, & Baker, 2003). Our final comments surround the challenges for research. Three key recommendations arise: first, establishing a recognized gold standard test that can be used and recognized by researchers in the field; second, investigating the use of these screening tools with changes in mood and circumstances over time; and third, improving the quality of research and the reporting of evidence from which accurate judgments about the efficacy of such screening tools can be established.

## Appendix

### *Reprinted Screening Measures*

---

#### **Suicide Checklist**

Inmate's name \_\_\_\_\_ Comis \_\_\_\_\_  
 Date of Incarceration \_\_\_\_\_ Date of Assessment \_\_\_\_\_  
 (yr/mo/day) \_\_\_\_\_ (yr/mo/day) \_\_\_\_\_  
 Age \_\_\_\_\_ First offence  Yes  No  
 Marital status  Single  Married/common-law  Separated/divorced

	VERY MUCH IN EVIDENCE	SOMEWHAT IN EVIDENCE	NOT IN EVIDENCE
<i>SYMPTOMS OF DEPRESSION:</i>			
Suicidal thoughts, hallucinations, or death fantasies	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Crying	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Depressed mood (sad, unhappy)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Expression of hopelessness or helplessness about the future	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Expressions of worthlessness	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Loss of energy, interest, or motivation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Loss of appetite or recent weight loss	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Neglect of personal appearance	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Disturbance in normal sleep patterns	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Loss of sexual desire	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Loss of enjoyment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	PRESENT	ABSENT
<i>PAST HISTORY</i>		
Recent suicide attempt of gesture (within the last year)	<input type="checkbox"/>	<input type="checkbox"/>
Past history of suicide gestures (non-life threatening)	<input type="checkbox"/>	<input type="checkbox"/>
Past history of serious suicide attempts (life-threatening)	<input type="checkbox"/>	<input type="checkbox"/>
Recent death of a loved one or divorce	<input type="checkbox"/>	<input type="checkbox"/>
History of previous psychiatric treatment (in-patient or out-patient)	<input type="checkbox"/>	<input type="checkbox"/>
History of aggressive, violent, or impulsive behaviour	<input type="checkbox"/>	<input type="checkbox"/>

---

Source: Arboleda-Florez & Holley (1988).

*Suicide Probability Scale*

Example Item	None or Little of the Time	Some of the Time	Good Part of the Time	Most/All of the Time
1. When...mad I throw things				
4. ...think bad things				
6. ...much I can do worthwhile				
10. ...people appreciate me				
14. ...make many changes in...life				
17. ...no one will miss me				
19. ...people expect too much				
21. ...not worth continuing to live				
23. ...no friends to count on				
26. I feel...close to my mother				
31. ...worry about money				
33. ...feel tired and listless				
35. When made...break things				
36. I can't be happy...				

Source: Western Psychological Services.

*Suicide Concerns of Offenders in Prison Environment (SCOPE)*

Name:.....			Sex: Male Female				
Are you on remand? .....Yes.....No.....			Age:.....				
PLEASE READ THE FOLLOWING STATEMENT AND CIRCLE THE RESPONSE ON THE RIGHT TO INDICATE IF YOU AGREE OR DISAGREE	PLEASE CIRCLE YOUR RESPONSE						
	Strongly Agree	Mildly Agree	Agree	Disagree	Mildly Disagree	Strongly disagree	
	1. I will not feel lonely in my room on my own	1	2	3	4	5	6
	2. If I had a job I would not commit crime	1	2	3	4	5	6
	3. If I were feeling suicidal I would speak to someone	1	2	3	4	5	6
4. I will speak to an officer when I have a problem	1	2	3	4	5	6	

(continued)

5. If I were on remand I would not feel stressed out	1	2	3	4	5	6
6. I do not think about harming myself	1	2	3	4	5	6
7. I do not feel suicidal when I receive bad news	1	2	3	4	5	6
8. I feel fine about coming into this establishment	1	2	3	4	5	6
9. If I had been arrested I would try and get in contact with my family	1	2	3	4	5	6
10. If I had been arrested I would say I was sorry	1	2	3	4	5	6
11. If I am nervous I do not lose my appetite	1	2	3	4	5	6
12. If I stole money for drugs I would feel like I had let myself down	1	2	3	4	5	6
13. I do not feel fed up	1	2	3	4	5	6
14. I think that everyone likes me	1	2	3	4	5	6
15. The day before I am due in court I do not think about the future	1	2	3	4	5	6
16. I enjoy everything	1	2	3	4	5	6

*(continued)*

17. I do not feel helpless	1	2	3	4	5	6
18. If I worry about things I sleep OK	1	2	3	4	5	6
19. If I were depressed I would talk to someone	1	2	3	4	5	6
20. My family support me	1	2	3	4	5	6
21. I do not think about how I can end my life	1	2	3	4	5	6
22. If I had a fight with a prisoner I would ask to see the governor	1	2	3	4	5	6
23. I can think straight when I am depressed	1	2	3	4	5	6
24. I feel like there is hope in my life	1	2	3	4	5	6
25. If I were depressed I would not think about harming myself	1	2	3	4	6	6
26. If I had a supportive family I would not kill myself	1	2	3	4	5	6
27. I always turn up in court	1	2	3	4	5	6

Source: Perry, 2005.

## *Suicide Probability Scale*

Indicator (scored as present/absent)

1. The offender may be suicidal
2. The offender has made a previous suicide attempt
3. The offender has undergone recent psychological/psychiatric intervention
4. The offender has experienced recent loss of a relative/spouse
5. The offender is presently experiencing major problems (i.e., legal)
6. The offender is currently under influence of alcohol/drugs
7. The offender shows signs of depression
8. The offender has expressed suicidal ideation
9. The offender has a suicide plan

Source: Wichmann et al., 2000.

## Notes

1. Where  $a$  = test positive and screen positive,  $b$  = test positive and screen negative,  $c$  = test negative and screen positive, and  $d$  = test negative and screen negative.

2. Psychological Abstracts (from January 1980 to June 2001); MEDLINE (Silver Platter online from January 1980 to June 2001); Embase (from January 1980 to June 2001); Database of Reviews of Effectiveness (DARE online; from January 1980 to November 2004); NHS Economic Evaluation Database (NHSEED; from January 1980 to November 2004); Health Technology Assessment (HTA; from January 1980 to November 2004); Social, Psychological, Educational and Criminological trials register (C2-SPECTR; from January 1980 to November 2004); Criminal Justice Abstracts (from January 1980 to November 2004); Criminal Justice Periodical Index (from January 1980 to June 2001); PsycINFO (from January 1980 to June 2001); and the National Criminal Justice Reference Service (NCJRS; January from 1980 to June 2001).

3. Her Majesty's Inspectorate of Prison Reports (from the Home Office Web site: <http://www.homeoffice.gov.uk>), assessment procedures and protocols (gathered from U.K. secure establishments), unpublished theses (from Dissertation Abstracts Online), the World Wide Web, reference lists from identified studies, and contact with experts in the field.

4. (screen\* or assess\* or evaluat\*)  
 (offender\* or criminal \* or prisoner\* or delinquent\*) and (juvenile\* or teen\* or adoles\* or young\*)  
 emotional adjustment or coping behaviour or (emotion\* and behaviour\*)  
 (suicid\* or self harm or self mutilation or self destructive behaviour or (emotion\* and behaviour\*)  
 (incarcerat\* or admit\* or remand\* or confin\* or lock\*) and (prison\* or jail\* or reformatory\* or  
 secure\* or (correction\* and (facilit\* or institut\*)))

## Acknowledgment

The authors would like to thank Gareth Johnson for his help and advice on the development of the search strategies for this review.



## References

- Alonzo, T. A., & Pepe, M. S. (1998). *Assessing the accuracy of a new diagnostic test when a gold standard does not exist* (UW Biostatistics Working Paper Series, No. 156). Retrieved from <http://www.bepress.com/uwbiostat/paper156>
- Arboleda-Florez, J., & Holley, H. L. (1988). Development of a suicide screening instrument for use in a remand setting. *Canadian Journal of Psychiatry, 33*, 595-598.
- Arboleda-Florez, J., & Holley, H. L. (1989). Predicting suicide behaviours in incarcerated settings. *Canadian Journal of Psychiatry, 34*, 668-674.
- Archer, R. P. (1989). Use of the MMPI with adolescents in forensic settings. *Forensic Reports, 2*, 65-87.
- Arensault-Lapierre, G., Kim, C., & Turecki, G. (2004). Psychiatric diagnoses in 3,275 suicides: A meta-analysis. *BMC Psychiatry, 4*, 1-11.
- Battle, A. O., Battle, M. V., & Tolley, E. A. (1993). Potential for suicide and aggression in delinquents at juvenile court in a southern city. *Suicide and Life Threatening Behaviour, 23*, 230-254.
- Birmingham, L., Gray, J., Mason, D., & Grubin, D. (2000). Mental illness at reception into prison. *Criminal Behaviour and Mental Health, 10*, 77-87.
- Blaauw, E., Kerhof, A. J. F. M., & Hayes, L. M. (2005). Demographic, criminal and psychiatric factors related to inmate suicide. *Suicide and Life Threatening Behaviour, 35*, 63-75.
- Blaauw, E., Kerhof, A. J. F. M., Winkel, F. W., & Sheridan, L. (2001). Identifying suicide risk in penal institutions in the Netherlands. *The British Journal of Forensic Practice, 3*, 22-48.
- Boothby, J. L., & Durham, T. W. (1999). Screening for depression in prisoners using the Beck Depression Inventory. *Criminal Justice and Behaviour, 26*, 107-125.
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, P. P., Glasziou, P. P., Irwig, L. M., et al. (2003). Standards for reporting of diagnostic accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *British Medical Journal, 326*, 41-44.
- Bureau of Justice Statistics. (2005). Prison statistics. Retrieved from <http://www.ojp.usdoj.gov/bjs/prisons.htm>
- Cole, D. A. (1989). Validation of the reasons for living inventory in general and delinquent adolescent samples. *Journal of Abnormal Child Psychology, 17*, 13-27.
- Cooke, D. J. (1995). Psychopathic disturbance in the Scottish prison population: Cross-cultural generalizability of the Hare Psychopathy Checklist. *Psychology, Crime, and Law, 2*, 101-118.
- Correia, K. M. (2000). Suicide assessment in a prison environment: A proposed protocol. *Criminal Justice and Behaviour, 27*, 581-599.
- Cull, J., & Gill, W. (1988). *Suicide Probability Scale (SPS) manual*. Los Angeles, CA: Western Psychological Services.
- Cullen, A., Long, C., & Lorincz, A. (1997). Rapid detection and typing in herpes simplex virus DNA in clinical specimens by the hybrid capture II signal amplification probe test. *Journal of Clinical Microbiology, 35*, 2275-2278.
- Daigle, M., Alarie, M., & Lefebvre, P. (1999). Problem suicide among female prisoners. *Forum on Corrections Research, 11*, 41-45.
- Defrancesco, J. J., Armstrong, S. S., & Russolillo, P. J. (1996). On the reliability of the youth self report. *Psychological Reports, 79*, 322.
- Department of Health. (1998). *Saving lives: Our healthier nation white paper*. London, UK: Department of Health.
- Earthrowl, M., & McCully, R. (2002). Screening new inmates in a female prison. *Journal of Forensic Psychiatry, 13*, 428-439.
- Esposito, C. L., & Clum, G. A. (1999). Specificity of depressive symptoms and suicidality in a juvenile delinquent population. *Journal of Psychopathology and Behavioural Assessment, 21*, 171-182.
- Falloon, S. (2000). *Suicide risk screening at Her Majesty's Prison, Leeds*. Unpublished report, Leeds.
- Faraone, S. V., & Tsuang, M. T. (1994). Measuring diagnostic accuracy in the absence of a gold standard. *American Journal of Psychiatry, 151*, 650-657.

- Fazel, S., Cartwright, J., Norman-Nott, A., & Hawton, K. (2008). Suicide in prisoners: A systematic review of risk factors. *Journal of Clinical Psychiatry, 69*, 1721-1732.
- Frühwald, S., Frottier, P., Benda, N., Eher, R., König, F., & Matsching, T. (2002). Psychosocial characteristics of jail and prison suicide victims. *Wiener Klinische Wochenschrift, 114*, 657-659.
- Grisso, T. R., Barnum, R., Fletcher, K. E., Cauffman, E., & Peuschold, D. (2001). Massachusetts Youth Screening Instrument for mental health needs of juvenile justice youths. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 541-548.
- Guo, B., Scott, A., & Bowker, S. (2003). *Suicide prevention strategies: Evidence from systematic reviews* (HTA Report No. 28). Edmonton, Canada: Alberta Heritage Foundation for Medical Research.
- Hayes, A. J., & Lever-Green, G. (2006). Developments in suicide prevention training for prison staff: STORM and beyond. *The Journal of Mental Health Workforce Development, 1*, 23-28.
- Huckaby, W. J., Kohler, M., Garner, E. H., & Steiner, H. (1998). A comparison between the Weinberger Adjustment Inventory and the Minnesota Multiphasic Personality Inventory with incarcerated adolescent males. *Child Psychiatry and Human Development, 28*, 273-285.
- Inch, H., Rowlands, P., & Soliman, R. (1995). Deliberate self-harm in a young offenders' institution. *Journal of Forensic Psychiatry, 6*, 161-171.
- Ivanoff, A., & Jang, J. S. (1991). The role of hopelessness and social desirability in predicting suicidal behaviour: A study of prison inmates. *Journal of Consulting and Clinical Psychology, 59*, 394-399.
- Ivanoff, A., & Jang, S. J. (1994). Prison suicidal behaviors interview. *Journal of Psychopathology and Behavioral Assessment, 16*, 1-13.
- Ivanoff, A., Smyth, N. J., Grochowski, S., & Jang, J. S. (1992). Problem solving and suicidality among prison inmates: Another look at state versus trait. *Journal of Consulting and Clinical Psychology, 60*, 970-973.
- Jenkins, R., Bhugra, D., Meltzer, H., Singleton, N., Bebbington, P., Brugha, T., et al. (2005). Psychiatric and social aspects of suicidal behaviour in prisons. *Psychological Medicine, 35*, 257-269.
- Kempton, T., & Forehand, R. (1992). Suicide attempts among juvenile delinquents: The contribution of mental health factors. *Behavioural Research Therapy, 30*, 537-541.
- Lader, D., Singleton, N., & Meltzer, H. (1998). *Psychiatric morbidity among young offenders in England and Wales*. London, UK: Office of National Statistics.
- Liebling, A. (1992). *Suicides in prisons*. London, UK: Routledge.
- Mills, J. F., & Kroner, D. G. (2000). *Depression, Hopelessness and Suicide Screening Form (DHS): User guide*. Unpublished Manuscript.
- Mitichison, S., Rix, K. J., Renvoize, E. B., & Schweiger, M. (1994). Recorded psychiatric morbidity in a large prison for male remanded and sentenced prisoners. *Medicine, Science, and the Law, 34*, 324-330.
- Perry, A. E. (2005). *Suicide and self-harm risk in offenders: The development of a new psychometric scale*. PhD thesis, University of York.
- Perry, A. E., & Marandos, R. (in press). Screening and assessment tools assessing risk of suicide and deliberate self-harm in adult offenders. Available from <http://www.campbellcollaboration.org/>
- Perry, A. E., & Olason, D. T. (2009). A new psychometric instrument assessing vulnerability to risk of suicide and self-harm behaviour in offenders: Suicide Concerns for Offenders in Prison Environment (SCOPE). *International Journal of Offender Therapy and Comparative Criminology, 53*, 385-400.
- Phillips, C. C., & Dahlstrom, W. G. (1997). Adapting the MMPI for use in assessing late adolescents in Trinidad and Tobago. *Adolescence, 32*, 887-896.
- Pratt, D., Piper, M., Appleby, L., Roger, W., & Shaw, J. (2006). Suicide in recently released prisoners: A population-based cohort study. *Lancet, 368*(9530), 119-123.
- Rohde, P. S., Seeley, J. R., & Mace, D. E. (1997). Correlates of suicidal behaviour in a juvenile detention population. *Suicide and Life Threatening Behaviour, 27*, 164-175.
- Sackett, D. L., & Haynes, R. B. (2002). The architecture of diagnostic research. In J. A. Knottnerus (Ed.), *The evidence base of clinical diagnosis* (pp. 19-38). London, UK: BMJ.
- Samaritans. (1998). *Exploring the taboo*. London, UK: Author.
- Shaw, J., Appleby, L., & Baker, D. (2003). Safer prisons—A national study of prison suicides 1999–2000 by the National Confidential Inquiry into Suicide and Homicide by People with Mental Illness. London, UK: Department of Health.

- Shaw, J., Baker, D., Hunt, I. M., Moloney, A., & Appleby, L. (2004). Suicide by prisoners. National clinical survey. *British Journal of Psychiatry, 184*, 263-267.
- Skegg, K. (2005). Self-harm. *Lancet, 28*, 1471-1483.
- Stein, A. L., & Yeager, C. A. (1993). Juvenile justice assessment instrument. *Juvenile and Family Court Journal, 44*, 91-102.
- Wichmann, C., Serin, R., & Motiuk, L. (2000). *Predicting suicide attempts among male offenders in federal penitentiaries* (Research Report R-91). Ottawa, Ontario: Research Branch, Correctional Service of Canada.