



Fourth DRAFT- February 2016

Harmful sexual behaviour in children: Evidence for identifying and assessing risk in children and young people who display harmful sexual behaviour.

Reference: CPH HSP ER (Evidence Reviews)

Authors:

Fiona Campbell¹, Andrew Booth¹, Evgenia Stepanova², Simon Hackett²,
Anthea Sutton¹, Kevin Hynes³, Jean Sanderson¹, Karen Rogstad⁴

¹ School of Health and Related Research (ScHARR)

University of Sheffield

Regent Court,

30 Regent Street,

Sheffield, S1 4DA,

² School of Applied Social Sciences (SASS)

Durham University

29 Old Elvet

Durham, DH1

³ Barnardo's

⁴ Sheffield Teaching Hospitals NHS Foundation Trust

Table of Contents

EXECUTIVE SUMMARY	5
Summary of Quantitative Data.....	5
Aim.....	5
Methods	5
Findings	5
Conclusion	6
Summary of Qualitative Data.....	7
LIST OF ABBREVIATIONS	10
GLOSSARY.....	12
EVIDENCE STATEMENTS	13
Quantitative Review	13
Qualitative Review	18
BACKGROUND.....	22
AIMS AND OBJECTIVES.....	23
Research Questions.....	23
METHODS	24
Identification of evidence.....	24
Results.....	29
DISCUSSION.....	55
Conclusion	58
A qualitative evidence synthesis of attitudes, barriers and facilitators to assessment of children and young people who display harmful sexual behaviour.....	61
AIMS AND BACKGROUND	61
METHODS	63
SUMMARY OF INCLUDED STUDIES	68
STUDY FINDINGS.....	78

General assessment tools specific to young people with Harmful Sexual Behaviour	87
Risk assessment tools specific to young people with HSB (i.e. that estimate risk of recidivism)	88
Generic assessment tools for children and youth at risk of harm or harming others.....	91
DISCUSSION	97
Statement of principal findings.....	97
Integration of quantitative and qualitative and expert evidence	Error! Bookmark not defined.
References	104
Appendix 1 - Excluded studies with reasons	108
Appendix 2 - Sample Search Strategy (from Ovid MEDLINE).....	115
Appendix 3 – Summary of assessment tools	118
Appendix 4 Evidence Tables – Quantitative Review.....	122

EXECUTIVE SUMMARY

Summary of Quantitative Data

Aim

The overall review question is as follows:

How effective are different models or tools in identifying the level of risk posed by, and address the needs of, children and young people who display harmful sexual behaviour (HSB)?

Within this main review question the review aimed to investigate and describe data on the effectiveness of the assessment measures and tools focusing on risk and need of children and young people who display harmful sexual behaviour.

Methods

We conducted a systematic search with multiple health and social care databases. The citations for all included references were added to identify the relevant studies. The initial search identified 4513 studies which were initially included in the process of sifting (3711 studies were left after duplicates have been removed). Additional papers were located through the references stated in previous reviews or mentioned in the annotated bibliographies of relevant risk assessment tool studies. After the abstracts were scrutinised to examine whether the papers showed evidence of implementing any of the risk assessment tools or psychometric tests for young sexual offenders, the number of studies was reduced to 322. Next, a number of studies were excluded due to unsuitable sample ages, or because of their focus on victims of sexual and physical abuse, intimate partner physical violence or adolescent groups with risk taking sexual behaviour (as opposed to HSB). The full-texts of 91 studies were assessed and 80 articles excluded. Finally, 11 studies were included.

Findings

We grouped assessment tools into two categories; those assessing behaviours and used to screen for HSB enabling differentiation between behaviours that might fall within normal developmental parameters and that those that are HSB. The second category of tools, were those used for the assessment of HSB behaviours and used in assessment, including the assessment of risk of recidivism. There is no quantitative data evaluating the quality of this

first group of tools (The Brook Traffic Light Tool and ...) indicating a gap in research evidence supporting practice. We included eleven studies reporting quantitative data on the validity of risk assessment tools for children and young people who display HSB. As only a small proportion of children and adolescents who display HSB commit future sexual offences, the limitation of this evidence is that it does not inform an understanding of the effectiveness of tools used in assessment for the majority of children who display HSB. Eleven included studies evaluated the predictive validity of risk assessment tools designed to identify adolescents who have committed sex offences and who are at risk of future sexual and/or nonsexual recidivism. There was data evaluating nine risk assessment tools; AIM2, adjusted AIM, J-SOAP II, ERASOR, J-SORRAT-II, SAVRY, YLS/CMI, PCL:YV and Static-99. Griffin et al (2008) in a study of moderate quality evaluated the predictive accuracy of AIM2. It was effective in predicting the risk of future sexual recidivism (strengths scale AUC = 0.94 CI 0.89 to 1, $p < 0.00001$, and the concerns scale AUC = 0.94 CI 0.98 to 1, $p < 0.00001$). This evidence is highly applicable as the tool was designed for use in the UK context and is currently widely in use. There was evidence from three studies to support the use of ERASOR in predicting sexual recidivism in male adult sex offenders; the AUC results were; (0.71 (95% CI 0.62 to 0.80) 0.72 (95% CI 0.61 to 0.83), 0.77 (95% CI 0.61 to 0.92) respectively which were all statistically significant ($p < 0.05$) however the results were not consistent. One study did not find it was able to accurately predict sexual recidivism. None of the studies explored their effectiveness in predicting recidivism with adolescent females, younger children or young people with learning difficulties. Meta-analysis of the included studies looking at sexual recidivism were: ERASOR (AUC 0.71 95%CI 0.65 to 0.77), J-SOAP (AUC 0.68 95% CI 0.56 to 0.84), J-SORRAT (AUC 0.61 95% CI 0.52 to 0.71)

Conclusion

The evidence regarding assessment tools, both those tools used to assess concerning behaviours and identify HSB and those tools used to assess children and young people who display HSB, is very limited. The AIM2 appears to show potential value as a tool that is accurate in predicting risk but is also designed for use within a UK context and to promote and facilitate holistic assessment and inter-agency working. There is also evidence that ERASOR and J-SOAP may assist in accurate prediction of risk. However, assessment of risk is only one part of an assessment of a child or adolescent with HSB and in order to ensure accurate assessment and appropriate support the assessment process needs to consider a wide range of other factors that can impact upon the utility and accuracy of the assessment tool. These are considered in the qualitative evidence review and in the integration of the qualitative and quantitative evidence.

Summary of Qualitative Data

Introduction

Assessing children and young people with HSB is considered important in a variety of contexts. It enables practitioners to understand the view their clients have about their situation and the significance of their sexual behaviour problems. Assessment helps to identify risks and barriers to change, as well as strengths, and young people's motivation to change. Assessment informs young people's suitability for intervention and their treatment needs (Griffin & Beech, 2004). However the literature reveals considerable inconsistency in what constitutes an assessment. As a consequence, those who use the assessment may exhibit substantive differences in how they view the assessment process. In summary assessment has been identified as having five key goals (Hackett, 2004):

1. **problem explanation:** understanding the sexual behaviour within the context of the individual young person's overall psychosexual, emotional and social functioning.
2. **risk formulation:** identifying features that are relevant to considering level of risk.
3. **risk management:** identifying the degree of control, restriction or supervision required to manage assessed levels of risk.
4. **intervention planning:** identifying areas where change is needed and how it can be achieved to support the young person to live a non-abusive lifestyle.
5. **evaluation:** assessing how change will be evaluated and progress measured.

This qualitative evidence synthesis seeks to complement an effectiveness review by examining existing published and unpublished qualitative research to establish what methods or components of assessment are viewed as acceptable or useful by children or adolescents who display harmful sexual behaviour, their parents or carers, health or social care professionals and health or social care managers and what considerations should be addressed when seeking to implement the assessment process.

Aim

The overall review question was:

What types of assessment are effective and acceptable for children and young people who display harmful sexual behaviour (HSB)?

Within this overall question the qualitative review component aimed to identify data on the assessment process from diverse stakeholder perspectives (i.e. young people, their family and carers, health and social care professionals and service managers).

Methods

We conducted specific searches across multiple health and social care databases. We pursued citations for all included studies in an attempt to identify related studies. We examined a larger subset of almost 1727 references (including duplicates) that had been coded as containing potential qualitative aspects. It was not possible in most cases to identify the presence of assessment within the titles and abstracts so the inclusion of most studies was based on citation chasing and examination of full-text.

Findings

We have identified 11 studies offering qualitative perspectives on the assessment of children and young people who display HSB. Nine of the studies were from the UK and two were from New Zealand. User perspectives of assessment were limited with most studies being conducted with either social care professions and youth offender teams or their managers. Assessment should be viewed as the first part of the therapeutic pathway and represents first contact for the adolescent and their family with social care agencies. Relationships between the adolescent and the professional delivering assessment are key. In addition the adolescent and their family should be able to observe that assessments are being used by professionals in order to help the young person, and in so doing increase the engagement and confidence of the family and young person in the process of assessment. A tension was identified between the requirements of criminal justice and social care systems with procedures of the former potentially delaying assessments. Delayed or incomplete assessments had a negative effect on the offender and family preventing passage to intervention and further social care provision. Social care professionals often framed delays using the language of risk and safety. Many tools, excluding AIM/AIM2 and ASSET lack qualitative investigation, and qualitative data for those tools is derived exclusively from programme evaluations. The absence of qualitative data from the United States is noticeable and may be attributed to an overwhelming preoccupation, in research and practice, with quantitative tools for forensic assessment, actuarial methods and prediction of recidivism.

Conclusion

Children and young people who display HSB are a very diverse group, whose needs vary, where the causes of the behaviours differ, where the social, cultural and environmental contexts are individual and where the behaviours displayed may be very different. The evidence to date focuses primarily on male adolescents with HSB that has resulted in a sexual offence. This limited evidence base was also apparent when considering the evidence for treatment effectiveness. We were not able to identify research evidence examining the effectiveness of tools in the assessment of HSB for younger children, girls, those with behaviours that may be harmful to the child or adolescent but do not constitute an offence and those with special educational.

This review has also identified an ongoing need to investigate the qualitative aspects of assessment, both generally and in relation to specific tools. Such research should examine both organisational and multi-agency aspects and perceptions of specific tools. The client perspective is a particular deficit from existing research. We identified no comparative data directly examining differential perspectives of different approaches to assessment.

The evidence we have identified, suggests that tools such as AIM2, ERASOR and J-SOAP are effective in identifying male adolescents who have committed sexual offenses and are at risk of future re-offending. The qualitative evidence suggests that AIM2 is useful in promoting multiagency working and is currently used by youth offending teams, social care agencies and within the voluntary sector.

LIST OF ABBREVIATIONS

AIM	Assessment Intervention Moving on
AIM2	Assessment Intervention Moving on (2 nd version)
ASO	Adolescents who have sexually offended
ASSET	[Youth Justice Board] - structured risk assessment tool used by all Youth Justice Services teams in England and Wales with young people in the criminal justice system
BERS-2-2	Behavioural and Emotional Rating Scale
CBT	Cognitive Behaviour Therapy
CSCS	Children's Social Care Services.
CTQ	Childhood Trauma Questionnaire
DSCQ	Day-Care Sexuality Questionnaire
ERASOR	Estimate of Risk of Adolescent Sexual Offence Recidivism
FFT	Functional Family Therapy
FTP	Family Treatment Program
FWI	Fight With Insight
G-MAP	Greater Manchester Adolescent Programme- a UK service for children and young people with HSB
HSB	Harmful Sexual Behaviour
IASO	Intrafamilial Adolescent Sex Offenders
ICU	Inventory of Callous-Unemotional Traits
JSO	Juvenile Sex Offenders
J-SOAP-II	The Juvenile Sex Offender Assessment Protocol-II
J-SORRAT-II	Juvenile Sexual Offence Recidivism Risk Assessment Tool-II
JRAS	Juvenile Risk Assessment Scale (JRAS)
LA	Local Authority
MACI	Millon Adolescent Clinical Inventory
MEGAJ	Multiplex Empirically Guided Inventory of Ecological Aggregates for Assessing Sexually Abusive Children and young people (Ages 19 and Under)

MMPI-A	Minnesota Multiphasic Personality Inventory-Adolescent
PCL:YV	[Hare] Psychopathy Checklist: Youth Version (PCL:YV).
PSRs	Pre-sentencing reports
SAVRY	Structured Assessment of Violent Risk in Youth (SAVRY)
SHB	Sexually Harmful Behaviour
SRM-SF	The Sociomoral Reflection Measure-Short Form
STFA	Situation, Thought, Feeling and Action
SW	Social Workers
YLS	Youth Level of Service Case Management Inventory
YOT	Youth Offending Team
YO-LSI	Young Offender Level of Service Inventory
YSBP	Youth with Sexual Behaviour Problems

GLOSSARY

Risk assessment – the process of estimating and evaluating risk. A probability calculation that a harmful behaviour or event will occur, which involves an assessment about the frequency of the behaviour/event, its likely impact and who it will affect (from Kemshall and Pritchard, 1996).

Risk factors – comprise static and dynamic factors. Static, historic risk factors (e.g. age at first offence, sex, offence history, health record etc.) do not change, whereas dynamic factors (e.g. drug use, traumatic events) are variable.

EVIDENCE STATEMENTS

Quantitative Review

Evidence for assessment tools designed to establish risk of reoffending in young people who have committed sexual offences.

There is a lack of empirical evidence evaluating the effectiveness of assessment tools designed to assess potentially harmful sexual behaviours in children and young people, and distinguish between sexual behaviours that fall within normal developmental parameters and those which may be harmful to the child themselves or others. (ES2.1)

During the search for relevant research evidence, it was apparent that some tools have been developed to assist professionals in identifying those children and young people with HSB, and to determine the most appropriate referrals, actions and treatment. These include the; Brook Traffic Light tool, and a continuum model of children and young people's sexual behaviours, patterns and cycles (Hackett 2010). These are referred to in the review as they are widely used in practice and potentially offer a valuable resource in identifying children with HSB.

There is evidence that supports the use of AIM2 as a tool to predict risk of sexual recidivism in adolescent male sex offenders and also adolescent male sex offenders with intellectual impairment. (ES2.2)

There is evidence from two small-scale retrospective recidivist studies (+,+) ^{1,2} that the AIM2 assessment tool was able to accurately predict sexual recidivism in adolescent males who had a history of HSB (total concerns score AUC= 0.98, CI 0.98 to 1.01, strengths scale AUC=0.94, CI 0.89 to 1, p<0.00001).¹ It also was found to be effective in predicting sexual recidivism in adolescent males with intellectual disabilities Strengths scale (AUC = 0.94 CI 0.89 to 1, p=<0.00001).²

¹Griffin et al 2008, [+], UK

²Griffen & Vettor, 2012 [+], UK

Applicability

This evidence is highly applicable to the UK context, the AIM framework was introduced as a tool for use within Greater Manchester and it adopted an approach that brought together the Framework for the Assessment of Children in Need and their Families (Department of Health 2000) and ASSET (Youth Justice Board 2000)

The evidence for the effectiveness of the J-SOAP-II tool to predict future sexual recidivism is contradictory. (ES2.3)

There is evidence from three nonrandomised, validation studies (-,-,+) ^{1,2,3} that the J-SOAP-II assessment tool, was able to accurately predict future sexual recidivism. None of these studies were prospective in design and assessment of risk was made from clinic records which reduced their quality rating. The area under the receiving operating characteristic (ROC) curve statistics (AUC) was 0.80,¹ 0.78 (95% CI 0.66 to 0.91)² $p < 0.01$ and 0.69 (95% CI 0.60 to 0.78) $p < 0.01$.³

However, in contrast, two nonrandomised validation studies (+,+) ^{4,5} found that the J-SOAP-II total score did not significantly predict sexual reoffending in adolescent sexual offenders who were discharged from a non-secure residential treatment program. Viljoen et al (2007) did identify higher AUCs for older youth (aged 16 to 18 years) at discharge than younger youth (aged 12 to 15 at discharge).

¹Prentky 2006, (-) USA

²Martinez et al 2007, (-) USA

³Rajlic and Gretton 2010, (+) Canada

⁴Viljoen et al 2007, (+) Canada

⁵Elkovitch et al 2008 (+) USA

Applicability

The findings of these studies have some applicability to the UK context, however; they were retrospective in design which limits the strength of the findings and their generalisability.

J-SOAP-II tool may be effective at predicting non sexual offences in adolescents but the results are mixed and may differ by age. (ES2.4)

There is evidence from two validation studies (+,-) ^{1,2} which found that the J-SOAP-II could significantly predict serious nonsexual offences (AUC 0.77 (95% CI 0.72 to (0.84) $p < 0.05$ and AUC 0.76 (95% CI 0.61 to 0.91) $p < 0.05$. In contrast one study (+) ³ found that total scores on the J-SOAP-II did not significantly predict reoffending of any type, however the J-SOAP-II was

significantly better at predicting serious nonsexual violent offences in older youth than in younger youth (OR 3.30).

¹Rajlic & Gretton 2010 (+) Canada

²Martinez et al 2007 (-) USA

³Viljoen et al 2007 (+) Canada

Applicability

None of the studies were undertaken in the UK which also limits their applicability to the UK context where assessment may be undertaken by different professional groups.

There is evidence that ERASOR (Estimate of Risk of Adolescent Sexual Offence Recidivism) may predict sexual recidivism in male adolescent sex offenders, however results are mixed. (ES2.5)

Sexual recidivism

Three validation studies (+,++,++)^{1,3,4} predicted sexual recidivism significantly better than chance; equivalent to medium effect sizes (Rice and Harris 2005). The AUC results were; (0.71 (95% CI 0.62 to 0.80) 0.72 (95% CI 0.61 to 0.83), 0.77 (95% CI 0.61 to 0.92) respectively which were all statistically significant (p<0.05). One study (+)² did not significantly predict future sexual reoffending following discharge (AUC 0.54 (95% CI 0.39 to 0.68)).

¹Rajlic and Gretton 2010 (+) Canada

²Viljoen et al 2009 (+) Canada

³Worling et al 2012 (++) USA

⁴Worling and Langton 2015 (++) USA

There is evidence that ERASOR may be able to predict future non sexual recidivism, but the effect is not consistent across all studies. (ES2.6)

One validation study (+)¹ found that the ERASOR (total score) could significantly predict future risk of future non sexual recidivism AUC 0.71 (95% 0.69 to 0.79) p<0.05. One study (++)³ found that ERASOR (total score) could significantly predict non sexual violent behaviours AUC 0.65 (0.53 to 0.76) p< 0.05 but not nonsexual nonviolent reoffending, However, the results were not consistent across the studies, with one study (+)² finding that ERASOR was not able to predict future non sexual reoffending

¹Rajlic and Gretton 2010 (+) Canada

²Viljoen et al 2009 (+) Canada

³Worling et al 2012 (++) USA

Applicability

Only one study was prospective in design which limits the strength of the findings and their generalisability. None of the studies were undertaken in the UK which also limits their applicability to the UK context where assessment may be undertaken by different professional groups.

There is no evidence that tools focusing on strengths (BERS-2) enhance the accuracy of ERASOR to predict sexual re-offending in adolescent males who have committed a sexual offence. (ES2.7)

One study¹ (++) examined the effectiveness of the BERS-2 tool in adding to the accuracy of the ERASOR in predicting future sexual and non-sexual re-offending. The BERS-2 tool consists of 5 scales. The BERS-2 Affective Strength scale (measuring the capacity for emotional intimacy) was found to be significantly predictive of desistance from sexual reoffending over the course of the follow-up period. With this scale, a higher score indicated greater protection from risk of reoffending. The AUC of 0.23 (0.09 to 0.37) indicated that 77% of those who desisted from subsequent sexual crimes had a higher score on the AS scale relative to those adolescents who reoffended. However, it did not have incremental validity over and above the ERASOR

With respect to nonsexual crimes the BERS-2 School Functioning (SF) scale (measuring aspects of competence with school) significantly predicted desistance from continued nonsexual reoffending. Seventy two per cent of those who desisted from nonsexual reoffending had higher scores on the SF scale (AUC of 0.28 (0.07 to 0.49)). None of the other BERS-2 scales were predictive of nonsexual recidivism.

Applicability

In the UK, the AIM2 tool is widely used. This incorporates an assessment of strengths. This evidence would suggest there is some value in assessing strengths but in conjunction with an assessment of risk.

¹Worling and Langton 2015 (++) USA

There is inconsistent evidence that the J-SORRAT-II (Juvenile Sexual Offence Recidivism Risk Assessment Tool – II) is able to predict future sexual or nonsexual recidivism amongst adolescent male sex offenders. (ES2.8)

One validation study (-)¹ by the developers of the J-SORRAT-II tool found that in 494 participants it was able to predict future sexual recidivism (AUC 0.64 (95% CI 0.57 to 0.71) $p < 0.05$) amongst adolescent male sex offenders. However, another study of 169 participants (+)² assessing future sexual aggression or nonsexual aggression found no effect (AUC 0.53 (95% CI 0.36 to 0.70) and AUC 0.56 (95% CI 0.45 to 0.66))

¹Epperson & Ralston 2009 (-) USA

²Viljoen et al 2007 (+) Canada

Applicability

The findings of these studies have some applicability to the UK context, as they were conducted in independent samples. However, only one was prospective in design which limits the strength of the findings and their generalisability. None of the studies were undertaken in the UK which also limits their applicability to the UK context where assessment may be undertaken by different professional groups.

There is no evidence that the SAVRY (Structured Assessment of Violence Risk in Youth) tool is able to predict future sexual recidivism in adolescent male sex offenders, but weak evidence that it may predict serious nonsexual violent offences in older youth. (ES2.9)

One validation study(+)¹ found no evidence that the SAVRY assessment tool could predict sexual or nonsexual offences following discharge from a treatment programme for adolescent male sex offenders. There is some evidence from the same study that the tool is able to predict a serious nonsexual violent offence in older youth.

¹Viljoen et al 2007 (+) Canada

Applicability

The findings of these studies have some applicability to the UK context, as they were conducted in independent samples. However, only one was prospective in design which limits the strength of the findings and their generalisability. None of the studies were undertaken in the UK which

also limits their applicability to the UK context where assessment may be undertaken by different professional groups.

There is no evidence that the YLS/CMI (Youth Level of Service/Case Management Inventory), PCL:YV (The Hare Psychopathy Checklist: Youth Version) and Static-99 tools are able to predict sexual reoffending although the YLS/CMI and PC:YV may predict violent re-offence in adolescent male sex offenders. (ES2.10)

One validation study (+)¹ evaluated the The YLS/CMI, PCL:YV and Static-99 tools. None of the total scores or professional ratings significantly predicted sexual reoffending. The YLS/CMI (total scores and professional ratings and the PCL:YV significantly predicted nonsexual violence (non-sexual violent offence), any violence (sexual or non-sexual offence) and any re-offence (any non-traffic offence). The Static-99 did not significantly predict reoffending of any type.

¹Viljoen et al 2009 (+) Canada

Applicability

These findings have limited application to the UK context, where the process of referral and treatment within the criminal justice system differs.

Qualitative Review

A pathway approach initiated by early assessment, with an iterative approach to treatment and on-going re-assessment, is perceived as most appropriate, and increases client confidence. (ES2.11)

Seven qualitative studies (+,+,+,+,+,+,-,++)^{1,2,3,4,5,6,7} identified the need for continuity between the assessment process and intervention. The assessment process should be “joined up” with subsequent therapeutic interventions. This increases client confidence in the value of the assessment process. It is also helpful to try to ensure some continuity between assessment and intervention. Assessment should be viewed as the first part of the treatment pathway with the potential to influence the subsequent engagement with a client. Particularly critical is the need for prompt assessment with delays or inadequate assessments leading to a risk to safety and a subsequent delay in accessing appropriate interventions or services. The process of assessment itself presents risk of harm by potentially stigmatising the young person and their family. The rapport that is developed during the process of assessment should be built up and

developed further during the intervention. Rapport and trust was identified as a key component of successful interventions.

¹Baker et al, 2005 (+), UK

²Belton et al, 2014 (+), UK

³Deacon, 2015 (+), UK

⁴Geary, 2007 (++) , New Zealand

⁵Geary et al, 2011 (++) , New Zealand

⁶Griffin & Beech, 2004 (-), UK

⁷Griffin et al, 2008 (++) , UK

Applicability

These findings are very relevant and generalizable to the UK context. Five of the studies were conducted in the UK and the findings are supported by high quality qualitative evidence.

Poor inter-agency working, especially between criminal justice and social care sectors is perceived to affect assessment, referral and treatment and result in worse outcomes. (ES2.12)

Two qualitative studies (+,-) ^{1,2} highlighted the challenges posed by inter-agency divisions. The Children Act (1989) was said to be unhelpful in not identifying adolescents who sexually abuse others as children 'in need'. Some social care practitioners felt that assessment was better undertaken by criminal justice staff, because 'society had already indicated needs for sanctions when individuals offend norms'. Social care practitioners may simply be attempting to pass on the challenges of assessment to someone else. However they also expressed concerns over a lack of sufficient training, or a perceived inability to draw appropriately on previously learnt skills. Nevertheless, the evidence highlights the underlying perception of a lack of necessary ability. This links with the further finding that "lack of skills in challenging the denial of abusers and carers" was a major concern for practitioners. This study further reports "how to assess risk of reoffending and, more generally, what pertinent questions to ask when undertaking a comprehensive assessment" as identified skills gaps.

¹Deacon, 2015 (+), UK

²Ladwa-Thomas & Sanders, 1999 (-), UK

Applicability

These findings are very relevant and generalizable to the UK context. The studies were conducted in the UK and the findings are supported by high quality qualitative evidence.

Lack of training, and a lack of clarity over roles, is perceived as a barrier to sharing of assessments and to ‘ownership’ of the assessment by all practitioners. (ES2.13)

Evidence from seven qualitative studies (+,+,+,+,+,+,-)1,2,3,4,5,6,7 identified a range of barriers to undertaking comprehensive assessment of young people with HSB. These included lack of training and skills, poor interagency working and limitations with assessment tools. This led to delays which were considered potentially harmful. The assessment process was depicted as a type of bottleneck preventing young people from gaining access to appropriate care. Assessment is seen by some to be the responsibility only of the Youth Offending Teams (YOTs), Reluctance was also expressed at sharing of assessments and at using the assessments performed by another. However, it should not be inferred that assessment is a static one-off process because this can result in a lack of professional ownership of the results of the assessment by subsequent professionals involved in handling the case.

¹Belton et al, 2014 (+), UK

²Deacon, 2015 (+), UK

³Geary, 2007 (++) , New Zealand

⁴Geary et al, 2011 (++) , New Zealand

⁵Hall, 2006 (+), UK

⁶Hall, 2010 (+), UK

⁷Ladwa-Thomas & Sanders, 1999 (-), UK

Applicability

These findings are very relevant and generalizable to the UK context. Five of the studies were conducted in the UK and the findings are supported by high quality qualitative evidence.

There is evidence that development of the AIM2 assessment framework and the ASSET tool is perceived by practitioners to have facilitated progress towards a more standardised approach. (ES2.14)

Evidence from five qualitative studies (+,-,+,+,+)1,2,3,4,5 revealed that two tools commonly used in the UK (AIM2 and ASSET) had both strengths and weaknesses. ASSET, while helping to focus thinking and encourage a holistic assessment of the young person, was also found to be poorly structured, time consuming and not always fit for purpose. With the AIM2 tool, social workers expressed a frustration with a lack of

training in its use, and some confusion as to its purpose and how the findings might be applied in practice.

¹Deacon, 2015 (+), UK

²Ladwa-Thomas & Sanders, 1999 (-), UK

³Griffin et al, 2008 (++) , UK

⁴Roberts et al 2001 (+), UK

⁵Baker et al, 2005 (+), UK

Applicability

These findings are very relevant and generalizable to the UK context. Five of the studies were conducted in the UK and the findings are supported by high quality qualitative evidence.

BACKGROUND

Numerous factors make it difficult to assess the scale of the problem of children and young people who display harmful sexual behaviour (HSB). Official statistics and existing research suggest children and young people account for a significant minority of all sexual abuse perpetrated in the UK (Hackett, 2014). Children and young people who display HSB are a very diverse group, whose needs vary, where the causes of the behaviours differ, where the social, cultural and environmental contexts are individual and where the behaviours displayed may be very different. Once identified, assessing the risks presented by children and young people, as well as their own vulnerabilities and needs has been a challenging task for professionals.

There is evidence from studies in adult sex offenders that evaluations based on unstructured professional judgement are less accurate than structured risk assessments (Andrew et al 2006, Hanson & Morton-Bourgon 2009). As a result, there is a need to identify which risk assessment approaches is most appropriate for use with adolescents and children. Over the last 30 years, a range of tools and assessment models have been proposed for this task. They vary in type (for example, between those which emphasise the use of clinical judgement as opposed to more actuarially based models), the nature of the risk factors included (for example, the use of static as opposed to dynamic risk factors), as well as in the strength of the empirical evidence supporting their use.

This evidence synthesis seeks to complement an effectiveness review of interventions for children and young people with HSB. This review seeks to analyse existing published and unpublished research to identify what risk assessment tools and tests are used with children and young people who have demonstrated behaviours that raise concerns of HSB or have displayed HSB. Such tools help professionals to identify: factors relevant to the development and persistence of sexually problematic and abusive behaviours; strategies to manage situations of risk; and appropriate interventions.

AIMS AND OBJECTIVES

Research Questions

How effective are different models or tools in identifying the level of risk posed by, and address the needs of, children and young people who display harmful sexual behaviour (HSB)?

METHODS

Identification of evidence

Searches were conducted in August 2015 using a range of multi-disciplinary bibliographic databases. Following the findings of the initial scoping search and in discussions with the NICE, a two stranded approach was applied to the searches, whereby a specific search naming particular assessment tools was conducted, followed by a more sensitive search using generic assessment terms. All references from the specific search were screened. The references from the sensitive search were screened using the “progressive fractions” technique, a method developed by SCHARR and utilised in previous systematic reviews (Booth et al. 2015).

Search terms were developed from the scoping search and in discussion with the NICE team. Thesaurus and free-text terms were utilised, relating to the population (children and young people who demonstrate harmful sexual behaviour) combined with terms relating to assessment. The specific search focused on named assessment tools or the term “tool*” (including synonyms) in the title or abstract. The sensitive search utilised generic assessment terms, such as measurement, identification, diagnosis. These were combined with the population terms using adjacency operators such as “SAME” where available to ensure relevance, for example terms appearing in the same sentence. All searches were limited to English Language, Humans, and the publication time span of 1990-present.

Databases Searched:

MEDLINE via Ovid 1946-March Week 4 2015
Ovid MEDLINE In-Process & Other Non-Indexed Citations March 26, 2015
Embase via Ovid 1974 to 2015 March 26
Cochrane Database of Systematic Reviews via The Cochrane Library: Issue 3 of 12, March 2015
Database of Abstracts of Reviews of Effect via The Cochrane Library: Issue 1 of 4, January 2015
Cochrane Central Register of Controlled Trials via The Cochrane Library: Issue 2 of 12, February 2015
Health Technology Assessment Database via The Cochrane Library : Issue 1 of 4, January 2015
NHS Economic Evaluation Database via The Cochrane Library: Issue 1 of 4, January

2015
Science Citation Index Expanded (SCI-EXPANDED) --1900-present and Social Sciences Citation Index (SSCI) --1956-present via Web of Science
Social Care Online 1980-March 2015
PsycINFO via Ovid 1806 to March Week 4 2015
Social Policy and Practice via OvidSP 201503
EPPI-Centre - Bibliomap (mostly pre-2011), Dopher (2006-March 2015), TRoPHI (2004-March 2015)
The Campbell Library 2004-2015 (Volume 11)

Inclusion of relevant evidence

Two reviewers (FC, ES) independently, and blind to the others results, sifted the results of the searching in order to identify studies for inclusion in the review. We used pre-defined criteria for population, risk assessment tool, study design and outcomes to determine inclusion in the review.

Participants

- Children and young people aged under 18 years who display harmful sexual behaviour. In this guideline, the term 'children' refers to children under 10 – the age of criminal responsibility in England. The term 'young people' refers to those aged 10 to 18 and includes those serving community sentences, those on remand and those serving custodial sentences.
- Children and young people up to the age of 25 who display harmful sexual behaviour and have special educational needs or a disability. This age extension is in light of the Children and Families Act 2014.
- Formal and informal caregivers of children and young people aged under 18 who display harmful sexual behaviours.

Types of activities and tools

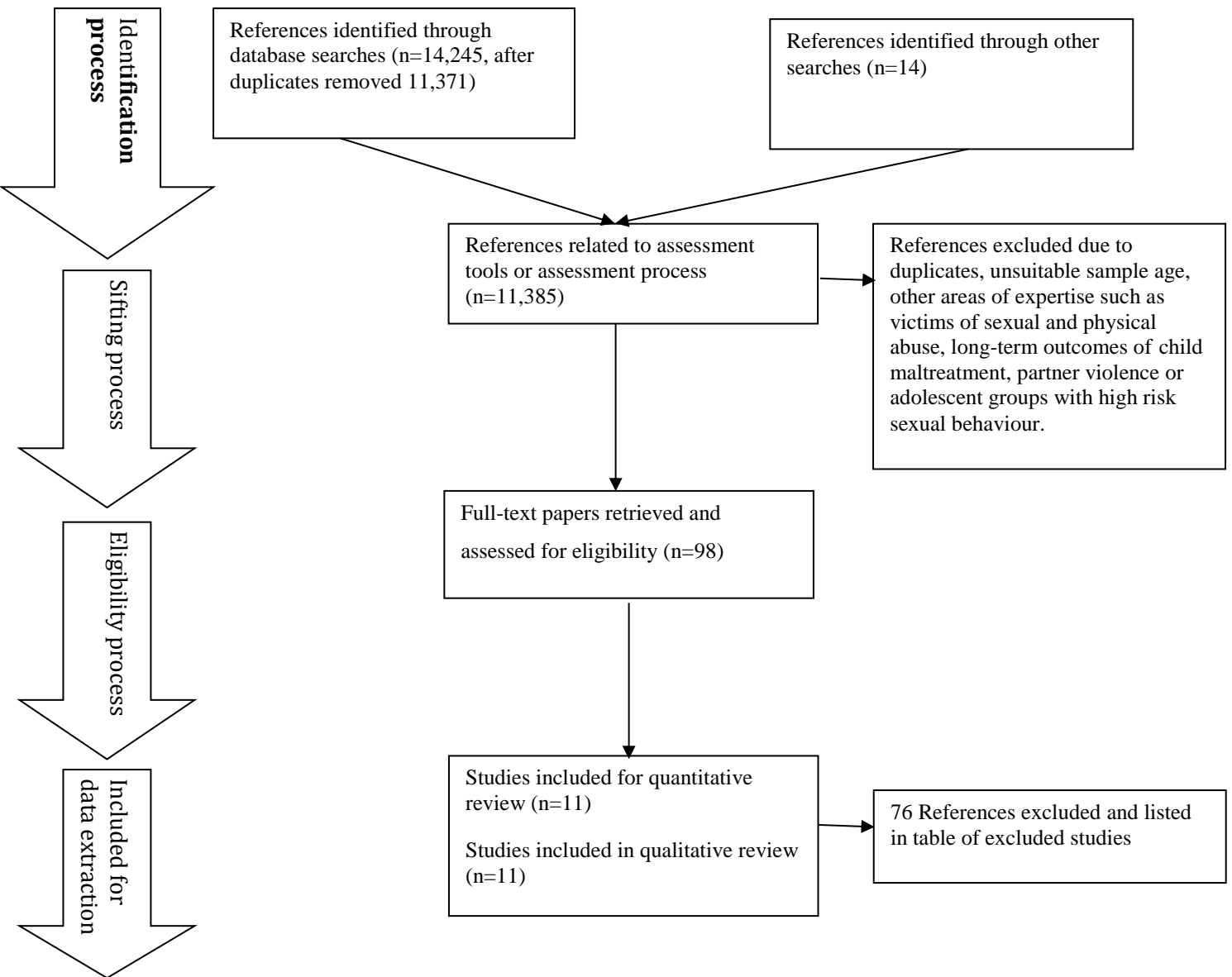
- Commissioning and partnership work (among the statutory, voluntary and private sectors) to identify, assess and help children and young people who display harmful sexual behaviour.
- Models or tools, including checklists that can distinguish between: normal behaviour, behaviour that needs to be assessed and monitored, and behaviour that needs a legal response and treatment.
- Assessment tools to identify the specific level of risk posed by children and young people who display harmful sexual behaviour and to identify how to address their needs.

Outcome Measures

- (Re)offence outcomes (sexual recidivism and non sexual offending/recidivism)
- Measures indicating training needs of staff using the tools
- Indicators of effective multi-agency working

We identified 11,371 potentially relevant studies from searching the electronic databases, and an additional 14 studies from a search of bibliographies of relevant reviews of the topic. Ninety eight full text copies of potentially relevant papers, or papers where the abstract did not provide sufficient information to ascertain whether the paper met the inclusion criteria, were retrieved. On further detailed reading of the 98 papers, 76 were excluded. The reasons for exclusion included: assessment tools focusing on adult perpetrators of sex offences; no test of validity and tools designed for the evaluation of risk and needs of children who have been the victim of abuse. Eleven studies were identified for inclusion in the quantitative review, and 11 were identified for inclusion in the qualitative review. (see flow diagram 1 for a description of the process of identifying studies for inclusion).

Figure one: Flow diagram showing the process of study identification



Methods of analysis/synthesis

Once identified and retrieved, data was extracted from the included studies independently by two reviewers (FC, ES). We used a piloted data extraction tool, designed in collaboration with topic experts within the review team (SH, KH). The data extracted can be found in appendix 1. This was then subject to a narrative synthesis. If there is sufficient data that is sufficiently homogenous, we shall undertake a meta-analyses. Studies were grouped on the basis of the type of assessment tools and tests.

Results

We identified two broad categories of assessment tools.

- Category 1: Behaviour assessment tools used to screen for HSB in order to identify behaviours that are outside of safe and healthy development and identify behaviours that need to be noted, further information gathered and the need for appropriate action considered.
- Category 2: Risk assessment tools specific to young people with HSB are tools used once a child or young person has displayed HSB that is considered to be outside of safe and healthy behaviours. They are used to identify appropriate interventions, and assess risk to the child or adolescent and others.

Category 1: General assessment tools specific to young people with HSB

We did not identify any quantitative studies for the assessment tools in category one. Table 1 shows the names of the tools we identified in the literature and the relevant reference. The 'Situation, Thought, Feeling and Action' (STFA) tool and Finkelhor's four preconditions model are tools that tend to be used in the assessment and treatment of HSB. They are therefore dissimilar to the Brook Traffic Light tool and the continuum model described by Hackett (2010) which are used to determine if a sexual behaviour is outside of the scope of normal behaviours and is HSB requiring further assessment, referral and/or treatment. We have however kept them within the text of the review but they are not included in the evidence statements.

Table 1: Summary of general assessment tools

Type of risk assessment	Reference	Setting
Brook Traffic light tool	Yamamoto & Kitan (2015)	UK
A continuum model of children and young people's sexual behaviours, patterns and cycles	Hackett (2010)	UK
Situation, Thought, Feeling and Action (STFA)	Durham (2006)	UK
Finkelhor's four preconditions model	Smallbone and Cale (2005) Durham (2006) Tutty (1991) Durham (2006)	UK

Summary Description of general assessment tools for young people with HSB.

The Brook Traffic Light Tool is a resource developed to help professionals identify, understand and respond to adolescent sexual behaviours. Used for children or adolescents at risk and in need of professional intervention (Yamamoto & Kitan, 2015).

Situation, Thought, Feeling and Action (STFA) is a cognitive behavioural chain that can be used to analyse the behaviours already committed by the child and young person, and also assist him in avoiding particular behaviours in future, by becoming more aware of the cues and triggers associated with the behaviour. The child or a young person is asked to analyse a situation that he might put himself in, and to consider the interacting process of the thoughts and feelings that led to him embarking on a particular action (Durham, 2006). **Patterns and Cycles** is another cognitive behavioural chain involving the interplay between the thoughts, feelings and physiological responses, inappropriate sexual fantasies, 'grooming behaviours' and thinking errors. This framework is often used in combination with 'The four steps and the four stops programme' (Durham, 2006).

A continuum model of children and young people's sexual behaviours (Hackett, 2010) is a framework which describes children and young people's sexual behaviour from normative behaviours through to highly deviant behaviours in order to assist in the assessment of behaviours on a developmental continuum.

Finkelhor's four preconditions model (including The Four Steps and Four Stops programme) is a framework to understand and prevent sexual offending. It explains child sexual abuse in terms of four elements that are necessary for an offense to occur: 1) an underlying motivation to sexually abuse; 2) overcoming internal inhibitors; 3) overcoming external inhibitors; and 4) overcoming the resistance of the child. It proposes a hierarchical model which included individual factors related to the victim, abuser and the family as well as social and cultural factors. It provides an adaptable and flexible framework which can accommodate new research to enhance understandings of why sexual abuse occurs. The model accounts for both intra and extra familial sexual abuse (Smallbone and Cale, 2005; Durham, 2006; Tutty, 1991). The Four Steps and Four Stops' programme is a development of Finkelhor's (1984) four preconditions of sexual abuse. The idea is to provide opportunities to prevent abuse by providing close but subtle supervision and monitoring. As the work progresses, the young person should become more aware of his risks and increasingly able to avoid such opportunities himself.

Category 2: Risk assessment tools specific to young people with HSB

One of the purposes of tools used in the assessment of children and young people with HSB is to provide an empirically based estimate of risk of future risk of future offending behaviour. This is to inform a range of decisions, such as: appropriate placements, treatment planning and resource allocation. Assessing risk for sexual recidivism and violent recidivism among young people is particularly challenging, as it is a period of enormous change and development. The difficulty of distinguishing between children and young people who are high and low risk is underscored by the finding that many young people who engage in deviant behaviours desist as they mature (Moffitt 1993). Another key factor that makes it difficult to assess risk for recidivism amongst adolescents is the absence of well-validated approaches to guide judgments. Given the well-documented limitations of unstructured clinical judgments (e.g. Hanson and Bussiere 1998), several adolescent risk assessment tools have been developed. Although the development of these tools is a significant step, there is currently inadequate evidence regarding their predictive validity.

Population characteristics

Eleven studies were identified for inclusion in the review (Elkovitch et al 2008, Epperson & Ralston 2009, Griffin et al 2008, Griffin & Vettor 2012, Martinez et al 2007, Prentky 2006, Rajlic & Gretton 2010, Viljoen et al 2007, Viljoen et al 2009, Worling et al 2012, Worling & Langton 2015). One study (Worling and Lanton 2015) reported the results of a further analysis on a subgroup of participants in a larger study (Worling et al 2012). The number of participants in the studies ranged from 60 to 822. Only one study (Prentky 2006) included girls, this study also included younger children with a mean age of 12.4 years (boys) and 12.0 (girls). The mean age of the adolescent males in the remaining studies ranged from 14.9 years to 18.8 years. All of the other studies only included young men, convicted and referred for treatment for committing a sexual offence. In three studies (Rajlic and Gretton 2010, Villjoen et al 2007, Viljoen et al 2009) the majority of participants had committed a sexual offence against a victim who was three or more years younger than themselves. Nine of the studies were conducted either in Canada or the USA and two conducted in the UK (Griffin et al 2008, Griffin & Vetto 2012) Six studies (Martinez et al 2007, Prentky 2006, Rajlic & Gretton 2010, Viljoen et al 2007, Viljoen et al 2009, Griffin & Vetto 2012, Griffin et al 2008)) reported the ethnicity of the participants. In one study (Martinez et al 2007) the majority of participants were of Hispanic origin (50%). However, in the other studies reporting ethnicity, the majority were Caucasian (See Table 2 for a summary of participant characteristics).

Table 2: Population Characteristics

Study	Tool tested	n	Setting	% male	Ethnicity	Mean Age Yrs (SD)	Nature of offence
Elkovitch et al (2008)	J-SOAP-II	166	USA	100	Not described	15.31 (1.51)	Convicted for a sexual offence and mandated to receive treatment
Epperson & Ralston (2009)	J-SORRAT-II	494	USA	100	Not described	Range 11-17 yrs	Convicted juvenile sex offenders
Griffin et al (2008)	AIM2	70	UK	100	White British: 80% African Caribbean: 1% Mixed race: 1% Other: 1% Unknown: 11%	15.5	Selected from cases referred for adolescents who have displayed sexually harmful behaviour. 87% had committed contact sexual offences.
Griffin & Vettor (2012)	AIM2 and adapted AIM	46	UK	100	White British: 76% Mixed white and black Caribbean: 2% Mixed white and Asian: 2% Asian Pakistani: 2% Asian other: 2% Not known: 7%	15.3 (1.4)	Selected from cases referred for adolescents who have displayed sexually harmful behaviour. 87% had committed contact sexual offences. 60% had committed indecent assault
Martinez et al (2007)	J-SOAP-II	60	USA	100	Hispanic: 50% African American: 28% Caucasian: 17% Other: 5%	14.9 (1.47)	Convicted of a sexual offence and admitted to a community based adolescent sex offender treatment programme
Prentky (2006)	J-SOAP-II	822	USA	81.1	Caucasian: 59% African-American: 11.6% Hispanic: 16.1% Other: 13.5%	12.39 (boys) 11.93 (girls)	Children who had begun to engage in sexually inappropriate and/or coercive acts with other children.
Rajlic & Gretton (2010)	J-SOAP-II ERASOR	286	Canada	100	Caucasian: 66% First Nations: 26% Asian, East Indian or Hispanic: 8%	15.8 (1.5)	91.8% had been convicted of a sexual offence, the remaining had a history of inappropriate sexual behaviour.. 26.2 offended against victims young than 5 years
Viljoen et al (2007)	J-SORRAT-II J-SOAP-II SAVRY	169	Canada	100	Caucasian: 83.4% African American: 4.7% Hispanic: 4.7% American Indian or Alaskan Native: 1.2%	15.37 (1.51)	Referred to a residential treatment programme. Majority had at least one index offence victim who was 3 or more years younger than the

					Mixed race: 2.4%		youth.
Viljoen et al (2009)	ERASOR YLS/CMI PCL:YV Static-99	193	Canada	100	Caucasian: 82.9% African American: 8.3% Hispanic: 1.6% American Indian or Alaskan Native:1.6% Unknown: 2.6%	15.26 (1.54)	Referred to a residential treatment programme. Majority had at least one index offence victim who was 3 or more years younger than the youth.
Worling et al (2012)	ERASOR	191	Canada	100	Not described	15.34 (1.53)	All convicted of and/or acknowledged criminal sexual behaviour.
Worling & Langton (2015)	ERASOR and BERS-2	81*	Canada		Not described	15.10 (1.53)	Convicted juvenile sex offenders

Quality Assessment of Included Studies

The included studies were critically appraised using the Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies (Wells et al 2009). Only two studies (Worling et al 2012, Worling and Langton 2015) were graded as high quality (++). These were the only studies that adopted a prospective design, with participants followed up following assessment using the risk assessment tool under investigation. All of the other included studies were retrospective in design. This raises the possibility that those using the assessment tool may become aware of the recidivism status of the participant. In three studies (Viljoen et al 2007, Viljoen et al 2009, Rajlic and Gretton 2010) those rating the participants were blind to the youth's subsequent charges. Martinez et al (2007) cautions that although blinding was attempted, contamination may have occurred and those using the assessment tools may have been aware of later reoffending. In three studies (Elkovitch et al 2008, Epperson & Ralston 2009, Prentky 2006) it is unclear if blinding occurred that ensured that those making the assessment were unaware of later recidivism status.

A further limitation of retrospective study designs that seek to validate risk assessment tools, is that they rely upon coding clinical file information. While these may be very comprehensive, there may be aspects of the risk assessment that would be more accurately assessed by a clinical interview with the participant rather than a reliance only on clinical records.

Duration of follow up was shortest (mean 3.66 years) for the studies of prospective design (Worling et al 2012, Worling and Langton 2015). Those that were retrospective ranged from a mean of 6 to 7 years. Duration of follow up was unclear in one study (Martinez et al 2007). There were no losses to follow up described in the two prospective studies (Worling et al 2012 and Worling & Langton 2015). The loss to follow up in the retrospective designed studies were reported in two studies (Rajlic and Gretton 2010, Viljoen et al 2007) and was 5% and 1.2% respectively. (See table 3 for a summary of the quality assessment grading)

Six studies (Griffin & Vettor 2012, Griffen et al 2008, Elkovitch 2008, Rajlic and Gretton 2010, Viljoen et al 2007, Viljoen et al 2009) were considered to be of moderate quality. Although these were retrospective studies, they attempted to blind raters to subsequent charges. Those excluded were described. Three studies (Epperson and Ralston 2009, Martinez et al 2007 and Prentky 2006) were rated as low quality, reflecting the lack of data to determine the methods used, and the possibility that the raters were aware of recidivism status as they coded.

Table 3: Quality Assessment (Newcastle Ottawa scale for cohort and case control studies)

	Selection				Comparability	Outcome			Score
	Representativeness of the exposed cohort	Selection of the non exposed cohort	Ascertainment of exposure (i.e. high risk of recidivism)	Demonstration that outcome of interest was not present at start of study	Comparability of cohorts on the basis of the design or analysis	Assessment of outcome	Was follow-up long enough for outcomes to occur	Adequacy of follow up of cohorts	
Elkovitch (2008) Prospective cohort study	Adolescent male sex offenders (USA)	Drawn from the same community as the exposed cohort	Graduate student rates who based assessment on comprehensive psychological file information. Had training and raters independently completed five practice files.	retrospective	Comparable for age, and all sexual offenders. Most had a victim much younger than themselves.	Independent records – state law enforcement sources, measured charges rather than conviction in case of possibility of dispositional bargaining.	Minimum 250 days. Mean 80.01 (SD 42) months	No statement	+
Epperson & Ralston (2009)	Adolescent male sex offenders (USA)	Drawn from the same community as the exposed cohort	ICC for coders	Not described	yes	Not described	Not described	No Statement	-
Griffin et al (2008)	Adolescent males with HSB (UK)	Drawn from the same community as the exposed cohort	Information from the adolescent's case file was used to complete the AIM2 and	Retrospective	Yes	Criminal histories were searched using the police national computer held	Mean 6 years, minimum 2 years	Excluded those where there was insufficient data recorded	+

			adapted AIM assessment models. Only information provided at the initial assessment was used			by the Ministry of Justice which provides a national database for England and Wales		at initial assessment	
Griffin & Vettor (2012)	Adolescent males with HSB (UK)	Drawn from the same community as the exposed cohort	Information from the adolescent's case file was used to complete the AIM2 and adapted AIM assessment models. Only information provided at the initial assessment was used	Retrospective	Yes	Criminal histories information was used to identify sexual re-offending and subsequent non-sexual offending was obtained from the police national computer and therefore these data had resulted from a charge or conviction. Also some cases, informed by up-to-date information from relevant professionals.	Mean 6 years (SD = 3.1, range 2-15.6 years)	Excluded those where there was insufficient data recorded at initial assessment	+

Martinez et al (2007)	Adolescent male sex offenders (USA) mainly Latin background – considered low risk	All adolescent males admitted to a community based treatment programme	Independent interrater reliability tested on a subset of cases. Treating clinicians scored.	Some potential contamination as outcome of interest may have been known before coding	yes	Not described	Not described	No statement	-
Prentky (2006)				Retrospective			Min: 6 m Max: 7y		-
Rajlic & Gretton (2010)	Adolescent male sex offenders referred for outpatient treatment (Canada)	Drawn from same community as exposed cohort	Not described	Retrospective coding – Risk measures were coded by trained research assistants blind to the criminal outcome information	yes	Limited to official criminal records – risk that sexual offences may have been coded as nonsexual offences.	Min: 12 m Max: 6.6 yrs	15 excluded due to missing data	+
Viljoen et al (2007)	Male adolescents	drawn from the same community as the exposed cohort	Risk assessment tools were coded on the basis of archival file information. This may result in lower scores and a more restricted range. Ideally	Raters blind to youth's subsequent charges. Based on file information and not clinical records	yes	Offending measured through juvenile and adult charges and by reviewing treatment records. Imperfect as it only included those cases	Average 7.24 years	2 excluded due to missing data	+

			need a combination of file information and interviews			that are detected. Should be multiple methods to assess reoffending such as self reported and parent reports in addition to official records.			
Viljoen et al (2009)	Adolescent male sex offenders	drawn from the same community as the exposed cohort	Secure record	Blind assessment Based on file information	yes	Reoffending was measured through official and unofficial sources, including law enforcement, probation and treatment records. Likely to underestimate reoffending.		None described	+
Worling et al (2012)	Adolescent male sex offenders	drawn from the same community as the exposed cohort	Coded by mental health professionals immediately following comprehensive	Prospective design	Also the adolescents were all subsequently involved in a specialised treatment	Relied on official documents regarding sexual and nonsexual		None described	++

			assessments.		focused on reducing their risk to reoffend. This may affect the accuracy of predictions of future risk	reoffending.			
Worling & Langton (2015)	Adolescent male sex offenders	drawn from the same community as the exposed cohort	Evaluators were trained. Evaluators rating following actual clinical assessments. These were comprehensive assessments that focused on multiple areas of functioning and involved multiple sources of information.	Prospective design – subset of 81 from above study		0.1 to 7.9 years (m = 3.66) sd 2.08		None described	++

Assessment tools

Hanson & Morton-Bourgon (2009) classified risk assessment procedures in one of the five following categories:

Empirical actuarial. These tools use explicit items determined in advance and explicit methods for combining the items into total scores. Both the items and the combination rules were selected on the basis of empirical evidence that linked them to recidivism. As well, these measures had tables that linked scores to expected recidivism rates

Mechanical. These tools use explicit items determined in advance and explicit methods for combining the items into a total score. They did not have a table that linked the total score to recidivism probabilities. The method of selecting and combining the items was based primarily on theory or literature reviews instead of direct analysis of specific data sets.

Adjusted actuarial. These evaluations were based on the total scores of an actuarial or mechanical tool except when the evaluator determined that there were factors external to the actuarial or mechanical scheme that justified overriding the obtained rating. The external factors were not specified in advance, and neither was the method of combining the external factors with the results of the actuarial or mechanical tool.

Structured professional judgment. Evaluators were given a structured list of risk factors determined in advance. The method of combining the factors into a total score was not specified in advance, and the overall evaluation of risk was left to the professional judgment of the evaluator.

Unstructured. Neither the risk factors nor the method of combining the factors into a total score was specified in advance. Risk assessments were based on individual case analysis, case conferences or professional experience.

We used this same scheme to categorise the included risk assessment tools

The eleven studies included in this review evaluated the following risk assessment tools: The Assessment, Intervention and Moving on Project (AIM2 and adjusted AIM), Estimate of Risk of Adolescent Sexual Offence Recidivism (ERASOR), Behavioural and Emotional Rating Scale (BERS-22), Juvenile Sex Offender Assessment Protocol –II (J-SOAP-II), Juvenile Sexual Offence Recidivism Risk Assessment Tool – II (J-SORRAT-II), Structured Assessment of Violence Risk in Youth (SAVRY), Youth Level of Service/Case Management Inventory (YLS/CMI), The Hare

Psychopathy Checklist: Youth Version (PCL:YV) and the Static-99. Four of these tools (AIM2, ERASOR, J-SOAP-II and J-SORRAT-II) were specifically designed to assess risk of reoffending in adolescent sex offenders. The other tools used in assessing risk have been developed for more generic types of offenders or assessing a broader range of offending behaviours. The list of tools, and their acronyms are shown in Table 4. More detail is given regarding each tool in the appendices.

Table 4: Tools evaluated in the included studies.

Tool	Acronym	Design purpose	Type of tool
The Assessment, Intervention and Moving on Project	AIM2	An integrated risks and strengths assessment to assist practitioners to identify risks and needs and assist them with their initial decision-making of young people who sexually abuse others	75 assessment items: 26 static concerns 6 static strengths 25 dynamic concerns 18 dynamic strengths
The Assessment, Intervention and Moving on Project	Adapted AIM	To be used in assessing and identifying the risk of recidivism in young people with intellectual difficulties who sexually offend.	101 assessment items, divided into concerns and strengths factors.
Behavioural and Emotional Rating Scale	BERS-22	This tool is self-report and measures strengths based and protective factors. It is designed to assess the personal strengths a child or adolescent possesses according to their own or an informant's perspective.	52 items, which form five scales tapping interpersonal strength, involvement with family, intrapersonal strength, school functioning and affective strength.
Estimate of Risk of Adolescent Sexual Offence Recidivism	ERASOR	Designed to assess risk of sexual violence among adolescents aged 12-18	
Juvenile Sex Offender Assessment Protocol	J-SOAP-II	Adolescents 12-18 designed to assess risk of sexual violence and antisocial behaviour	28 item checklist
Juvenile Sexual Offence Recidivism Risk Assessment Tool - II	J-SORRAT-II	Adolescents 12-18 designed to assess risk of sexual recidivism	12 item actuarial tool
Structured Assessment of Violence Risk in Youth	SAVRY	Adolescents 12-18 designed to assess risk of general (not specifically sexual) violent recidivism.	24 items to assist
Youth Level of Service/Case	YLS/CMI	Designed to assess risk	42 risk/need factors

Management Inventory		and needs and assist in development of a case management plan	
The Hare Psychopathy Checklist: Youth Version	PCL:YV	Adapted adult tool designed to measure psychopathic traits	20 item rating scale.
Static-99		Adult tool	10 item actuarial tool

FINDINGS

Eleven studies were identified and included in the quantitative review. Each explored the ability of risk assessment tools to predict sexual recidivism and some reported on their ability to accurately predict nonsexual recidivism. The population in whom these tools are used are those young people with HSB who have committed a sexual offence. This represents only a small proportion of children and young people with HSB and is a limitation of the evidence available. The outcomes are described narratively and where there is sufficient data, we have performed a meta-analysis. (See table 6 for a summary of the results)

Meta-analysis was performed in the R statistical software package using the metafor library. Data was synthesised using both fixed and random effects models. For the fixed effects meta-analyses inverse variance weight was used. For the random effects meta-analyses the DerSimonian-Laird estimator was used.

Where available, standard errors were calculated from the provided confidence intervals. For one of the tools (J-SOAP) not all studies provided confidence intervals for the AUC values. Two analyses were performed; the first excluding studies with unknown SE and the second using a conservative estimate of 0.1 for the missing values.

Recidivism

All of the included studies measured the validity of the risk assessment tool by recidivism rates. The duration of follow up ranged from a mean of 3.66 years to 7 years. All of the participants had participated in an intervention programme. This may have influenced the rates of sexual recidivism which were low across the studies (8.3 – 14.7%). The studies were also limited in that they relied on reported measures of recidivism and as many sexual offences are never reported, it is possible that the recidivism rates are therefore underestimates.

AIM2 and Adapted AIM

Sexual recidivism

Two included studies (Griffin et al 2008, Griffin & Vettor 2012) assessed accuracy of the AIM2 tool to predict sexual re-offending in a UK sample of adolescents. Griffin and Vettor (2012) tested both the AIM2 and the adapted AIM assessment for adolescents with intellectual difficulties. Adolescents with intellectual difficulties were defined as those with an IQ of 70 or below and impaired social functioning. Griffin et al (2008) examined 70 retrospective case files for young males of mainstream educational ability aged 12-18 years. It identified a subset of strengths and concerns items. Both the concerns scale items and the strengths scale made an independent contribution to risk prediction (area under the curve (AUC)= 0.98, CI =0.98 to 1.01 and AUC =0.94, CI 0.89 to 1.00). In 46 adolescents with intellectual difficulties the AIM2 concerns scale and the concerns scale for the adapted AIM assessment predicted sexual re-offending with approximately equal accuracy, yielding an AUC = 0.78 (CI: 0.63 to 0.94) and AUC = 0.78 (CI 0.62 to 0.94) respectively, indicating a large effect size. While the concerns scales used within these assessments were able to predict risk independently, the use of a combined strengths and concerns score on the AIM2 assessment was a slightly better predictor of sexual re-offending, although this difference was small (AUC = 0.79 (CI: 0.62 to 0.96)).

J-SOAP -II

Sexual recidivism

Five included studies (Martinez et al 2007, Prentky et al 2006, Rajlic and Gretton 2010, Viljoen et al 2007 and Elkovitch et al 2008) examined the predictive validity of the J-SOAP-II assessment tool. One of these (Elkovitch et al 2008) used both the J-SOAP and SAVRY tool to inform the clinical judgment of risk of reoffending. One study (Prentky 2006) found high accuracy for the J-SOAP-II total score in predicting sexual recidivism among their population. The area under the receiving operating characteristic (ROC) curve statistics (AUC) were 0.82 for preadolescents and 0.80 for adolescents (Prentky 2006). Two studies (Martinez et al 2007 and Rajlic and Gretton 2010) reported a moderate level of predictive accuracy 0.78 (95% CI 0.66 to 0.91) $p < 0.01$ and 0.69 (95% CI 0.60 to 0.78) $p < 0.01$. However, in contrast, two studies (Viljoen et al 2007 and Elkovitch et al 2008) found that the J-SOAP-II total score did not significantly predict sexual reoffending in adolescent sexual offenders who were discharged from a non-secure residential treatment program. Viljoen et al (2007) did identify higher AUCs for older youth (aged 16 to 18 years) at discharge than younger youth (aged 12 to 15 at discharge). The pooled effects of the three studies in predicting sexual recidivism (Viljoen et al 2007, Rajlic &

Gretton 2010, Martinez et al 2007) using a random effects model was AUC = 0.68 (95% CI 0.56 to 0.80) There was a high level of heterogeneity in this analysis (I^2 67) reflecting the differences in findings between studies. (see figure 2)

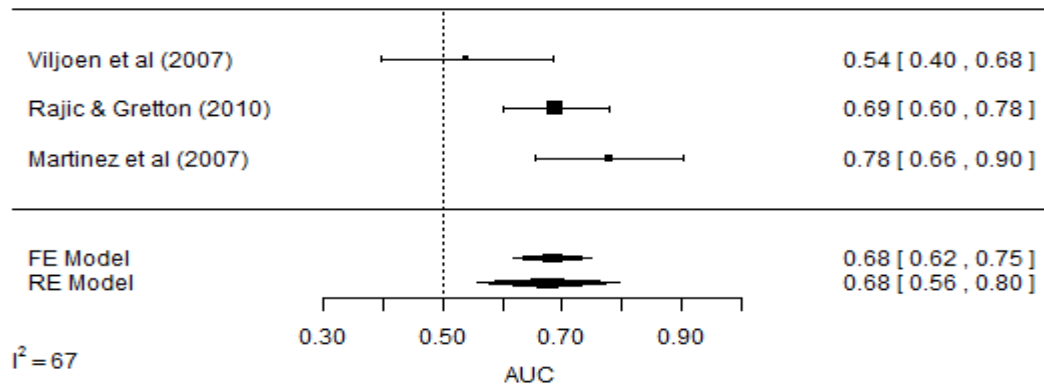


Figure 2: J-SOAP – Prediction of sexual recidivism

Non sexual recidivism

Rajlic & Gretton (2010) and Martinez et al (2007) found that the J-SOAP-II could significantly predict serious nonsexual offences (AUC 0.77 (95% CI 0.72 to (0.84) $p < 0.05$ and AUC 0.76 (95% CI 0.61 to 0.91) $p < 0.05$. In contrast Viljoen et al (2007) found that total scores on the J-SOAP-II did not significantly predict reoffending of any type, however the J-SOAP-II was significantly better at predicting serious nonsexual violent offences in older youth than in younger youth (OR 3.30). The pooled effects of the three studies (Viljoen et al 2007, Rajlic & Gretton 2010, Martinez et al 2007) in predicting non-sexual recidivism using a random effects model was AUC = 0.72 (95% CI 0.67 to 0.77) There was a high level of heterogeneity in this analysis (I^2 83) reflecting the differences in findings between studies. (see figure 3)

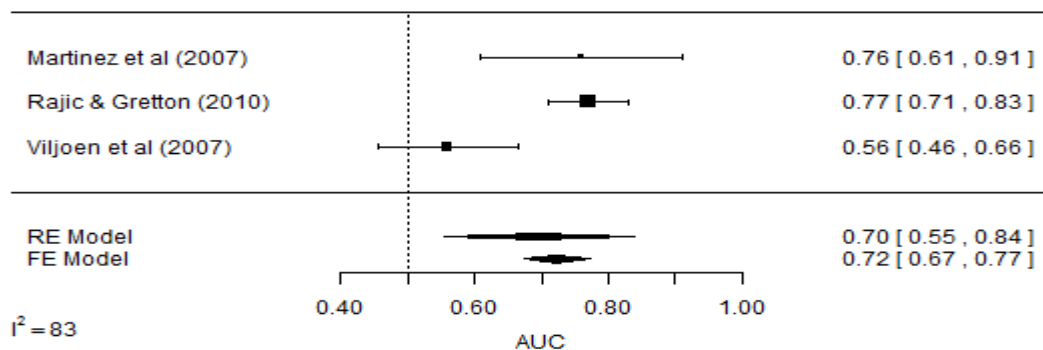


Figure 3: Prediction of non-sexual recidivism

ERASOR

Sexual recidivism

Four included studies (Rajlic and Gretton 2010, Viljoen et al 2009, Worling et al 2012, Worling and Langon 2015) examined the predictive validity of the ERASOR (total score). Three of the studies (Rajlic and Gretton 2010, Worling et al 2012, Worling and Langon 2015) predicted sexual recidivism significantly better than chance; equivalent to medium effect sizes (Rice and Harris 2005). The AUC results were; (0.71 (95% CI 0.62 to 0.80) 0.72 (95% CI 0.61 to 0.83), 0.77 (95% CI 0.61 to 0.92) respectively which were all statistically significant ($p < 0.05$). One study (Viljoen et al 2007) did not significantly predict future sexual reoffending following discharge (AUC 0.54 (95% CI 0.39 to 0.68)). The pooled effects of the four studies in predicting sexual recidivism (Rajlic and Gretton 2010, Viljoen et al 2009, Worling et al 2012, Worling and Langon 2015) using a random effects model was AUC = 0.71 (95% CI 0.6 to 0.77). There was no

evidence of heterogeneity in this analysis ($I^2=0$) (see figure 4)

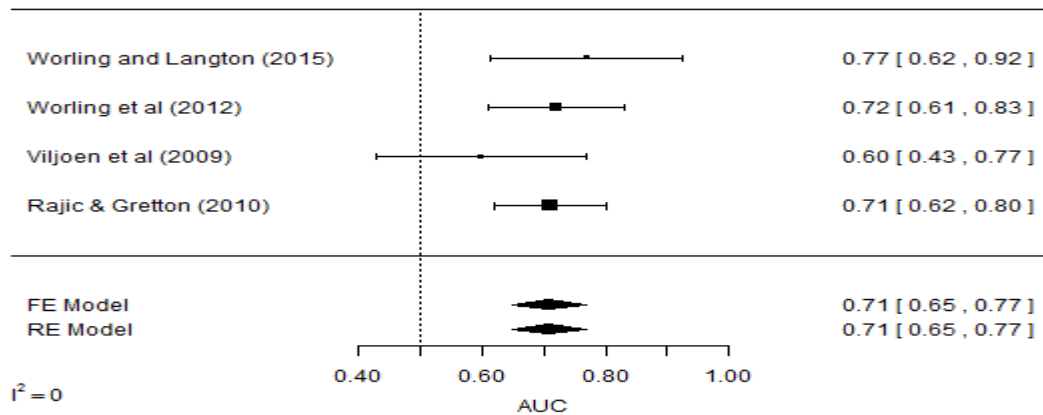


Figure 4: ERASOR prediction of sexual recidivism

Non sexual recidivism

Three studies (Rajlic & Gretton 2010, Viljoen et al 2009, Worling et al 2012) measured nonsexual recidivism. One of these (Rajlic & Gretton 2010) found that the ERASOR (total score) could significantly predict future risk of future non sexual recidivism: AUC 0.71 (95% 0.69 to 0.79) $p < 0.05$. One study (Worling et al 2012) found that ERASOR (total score) could significantly predict non sexual violent behaviours: AUC 0.65 (0.53 to 0.76) $p < 0.05$ but not nonsexual nonviolent reoffending, However, the results were not consistent across the studies, with the study by Viljoen et al (2007) finding that ERASOR was not able to predict future non sexual reoffending. The pooled effects of the four studies in predicting non-sexual recidivism (Rajlic and Gretton 2010, Viljoen et al 2009, Worling et al 2012, Worling and Langon 2015) using a random effects model was AUC = 0.65 (95% CI 0.57 to 0.73). There was moderate heterogeneity in this analysis ($I^2=48$) (see figure 5)

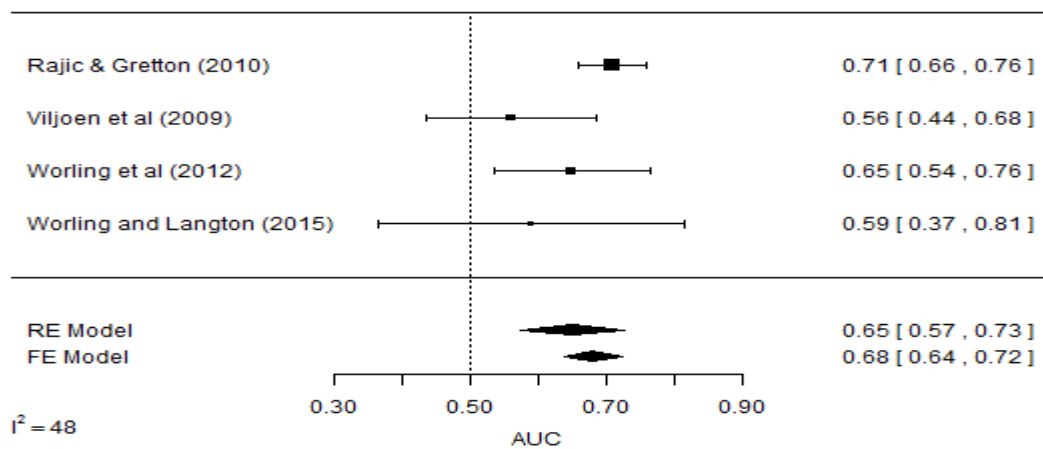


Figure 5: ERASOR – Prediction of non-sexual recidivism

J-SORRAT-II

Sexual and nonsexual recidivism

Two studies (Viljoen et al 2007 and Epperson & Ralston 2009) evaluated the predictive validity of the J-SORRAT-II tool. Viljoen et al (2007) found that total scores on the J-SORRAT-II did not significantly predict sexual aggression or nonsexual aggression (AUC 0.53 (95% CI 0.36 to 0.70) and AUC 0.56 (95% CI 0.45 to 0.66) respectively. However, a validation study by Epperson & Ralston (2009) did show a small but statistically significant chance of predicting sexual recidivism when using the J-SORRAT-II tool (AUC 0.64 (95% CI 0.57 to 0.71) $p < 0.05$). The pooled effects of two studies (Viljoen et al 2007 and Epperson & Ralston 2009) in predicting sexual using a random effects model was AUC = 0.61 (95% CI 0.52 to 0.71) and . There was some heterogeneity in this analysis ($I^2 = 27$) (see figure 6)

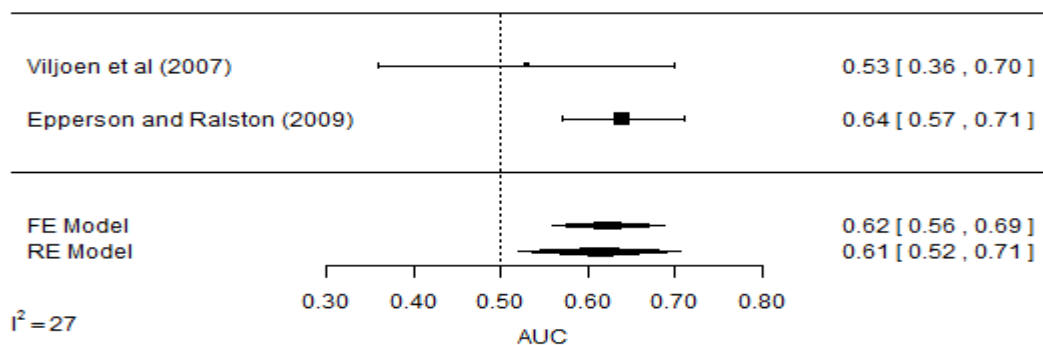


Figure 6: J-SORATT – Prediction of sexual recidivism

SAVRY

Sexual and nonsexual recidivism

One study (Viljoen et al 2007) evaluated the predictive validity of reoffending. SAVRY total scores did not significantly predict sexual or nonsexual offences following discharge. It did however predict serious nonsexual violent offences (AUC 0.69 (95% CI 0.56 to 0.81). Structured professional judgement, using SAVRY to assess an adolescent’s risk for violence (judged as low, moderate and high categories) did not significantly predict reoffending of any type. The SAVRY was however, better at predicting serious nonsexual violence offences in older youth (OR 3.05)

YLS/CMI, PCL:YV and Static-99

Sexual and nonsexual recidivism

The YLS/CMI, PCL:YV and Static-99 tools were evaluated in one study (Viljoen et al 2009). None of the total scores or professional ratings significantly predicted sexual reoffending. The YLS/CMI (total scores and professional ratings and the PCL:YV significantly predicted nonsexual violence (non-sexual violent offence), any violence (sexual or non-sexual offence) and any re-offence (any non-traffic offence). The Static-99 did not significantly predict reoffending of any type.

BERS-2

Sexual and nonsexual recidivism

There are five scales that are derived from the 52 items on the BERS-2 including; Interpersonal Strength (IS) scale, Family Involvement scale (FI), Intrapersonal Strength (IsS), School Functioning scale (SF), Affective Strength (AS). The BERS-2 Affective Strength scale (measuring the capacity for emotional intimacy) was found to be significantly predictive of desistance from sexual reoffending over the course of the follow-up period. With this scale, a higher score indicated greater protection from risk of reoffending. The AUC of 0.23 (0.09 to 0.37) indicated that 77% of those who desisted from subsequent sexual crimes had a higher score on the AS scale relative to those adolescents who reoffended.

With respect to nonsexual crimes the BERS-2 School Functioning (SF) scale (measuring aspects of competence with school) significantly predicted desistance from continued nonsexual reoffending. Seventy two per cent of those who desisted from nonsexual reoffending had higher scores on the SF scale (AUC of 0.28 (0.07 to 0.49)). None of the other BERS-2 scales were predictive of nonsexual recidivism.

Table 6: Outcomes of Predictive Validity

Study	Tool	n	Follow-up Mean (SD)	AUC (95% CI)	AUC* (95% CI)		ICC	Total Recidivism rate Sexual or non sexual Sexual offence Non sexual offence
				Sexual recidivism	Non sexual Recidivism			
Elkovitch et al (2008)	J-SOAP-II and SAVRY	161	6.6 (3.5)	Sexual violence 0.44	Nonsexual violence 0.58	Any violence 0.53	NR	14/ 161(8.4 %) 34/ 161(20.5%) 17/161 (10.2%)
Griffin et al (2008)	AIM2 concerns scale strengths scale	70	6	0.98 (0.98 to 1.01)*** 0.94 (0.89 to 1.00)***	NR	NR	52% of assessment items: >0.75 29% of items: 0.6 - 0.75 3% of items: <0.6	NR
Griffin & Vettor (2012)	AIM2 concerns scale strengths scale combined adapted AIM concerns scale strengths scale combined	46	6 (3.1)	0.78 (0.62 to 0.94) ^a 0.29 (0.09 to 0.49) ^b 0.79 (0.62 to 0.96) ^c 0.78 (0.63 to 0.94) ^d 0.51 (0.28 to 0.73) ^e	NR	NR		NR
Martinez et al (2007)	J-SOAP-II (TS)	60	Not described	0.78 (0.66 to 0.91)*	0.76 (0.61 to 0.91)*	NR	NR	NR

Prentky (2006)	J-SOAP-II	822	7 (R)	Preadolescence 0.82* Adolescence 0.80*	NR	NR	0.73 to 0.78	117/797 (14.7%) 28/797 (3.5%) 226/797
Rajlic & Gretton (2010)	J-SOAP-II (TS) ERASOR (TS)	286	6.6	0.69 (0.60 to 0.78)* 0.71 (0.62 to 0.80)*	0.77 (0.72 to 0.84)* 0.71 (0.69 to 0.79)*	NR	0.80 to 0.94 0.75 to 0.91	27/286 (9.4%) 124/286 (43.4%) 97/286 (33.9%)
Ralston & Epperson (2014)	J-SORRAT-II	494	6.58 (3.59)	0.64 (0.57 to 0.71)	NR	NR	0.96 (student coders) 0.91 (evaluators)	NR
Viljoen et al (2007)	J-SORRAT-II J-SOAP-II SAVRY (total scores) SAVRY (SPR)	169	6.6 (3.5)	0.53 (0.36 to 0.70) 0.54 (0.39 to 0.68) 0.53 (0.38 to 0.67) 0.51 (0.35 to 0.66)	Nonsexual violent 0.56 (0.45 to 0.66) 0.56 (0.45 to 0.66) 0.58 (0.48 to 0.68) 0.51 (0.40 to 0.62)	Any reoffence 0.54 (0.45 to 0.63) 0.56 (0.47 to 0.65) 0.58 (0.49 to 0.67) 0.50 (0.42 to 0.59)	0.89 0.84 0.91	14/169(8.3%) 21/169(12.7%) 71/169 (42.8%) any offence
Viljoen et al (2009)	ERASOR (TS) ERASOR (SPR) YLS/CMI (TS) YLS/CMI (SPR) PCL:YV Static-99	193	7.24	0.60 (0.43 to 0.77) 0.64 (0.49 to 0.79) 0.55 (0.41 to 0.70) 0.58 (0.43 to 0.73) 0.49 (0.33 to 0.65) 0.60 (0.44 to 0.76)	Nonsexual violent 0.56 (0.44 to 0.69) 0.54 (0.43 to 0.66) 0.68 (0.56 to 0.80)** 0.72 (0.61 to 0.82))** 0.71 (0.60 to 0.82)** 0.54 (0.42 to 0.66)	Any reoffence 0.53 (0.45 to 0.62) 0.50 (0.4 to 0.58) 0.66 (0.58 to 0.74)*** 0.60 (0.52 to 0.68)** 0.63 (0.55 to 0.71)** 0.52 (0.44 to 0.60)	ERASOR – 0.75 YLS-0.45	16/193 (8.3%) 25/193 (13%)
Worling et al (2012)	ERASOR - CJ ERASOR -TS ERASOR - sum	191	3.66 SD 2.08	0.61 (0.48 to 0.74) 0.72 (0.61 to 0.83)* 0.73 (0.63 to 0.84)*	Non sexual violent 0.61 (0.50 to 0.73) 0.65 (0.53 to 0.76)* 0.64 (0.52 to 0.76)	Nonsexual nonviolent reoffending 0.57 (0.47 to 0.68) 0.65 (0.41 to 0.90) 0.61 (0.41 to 0.82)		18/191 (9.4 %) 35/191 (18.32%) 26/191 (13.61%)
Worling and		81	3.66		Nonsexual			

Langton (2015)			SD 2.08		recidivism			
	ERASOR rating			0.63 (0.42 to 0.85)	0.59 (0.36 to 0.82)	NR		
	ERASOR sum			0.77 (0.61 to 0.92)*	0.59 (0.37 to 0.82)			
	BERS-2 IS			0.52 (0.36 to 0.67)	0.48 (0.30 to 0.67)			
	BERS-2 FI			0.38 (0.22 to 0.54)	0.37 (0.19 to 0.55)			
	BERS-2 IaS			0.50 (0.32 to 0.68)	0.50 (0.30 to 0.70)			7/81 (8.6%)
	BERS-2 SF			0.51 (0.28 to 0.74)	0.28 (0.07 to 0.49)*			9/81 (11.1 %)
	BERS-2 AS			0.23 (0.09 to 0.37)*	0.42 (0.20 to 0.63)			

*p<0.05 **p<0.01, ***p<0.0001, a: p=0.01, b: p=0.05, c: p=0.01, d: p=0.01, e: p=0.97 (NS)

TS=total score, SPR= structured professional rating BERS-2 = Behavioural and Emotional Rating Scale; ERASOR = +, ; ERASOR sum = ERASOR sum of risk factors rated present; BERS-2 Is = BERS-2 Interpersonal Strength; BERS-2 FI = BERS-2 Family Involvement; BERS-2 IaS = BERS-2 Intrapersonal Strength; BERS-2 SF = BERS-2 School Functioning; BERS-2 AS = BERS-2 Affective Strength. YLS/CMI = Youth Level of Service/Case Management Inventory. CL clinical judgment; TS total score

AUC = ROC analyses were used to evaluate the ability of risk assessment tools to predict youths treatment behaviour and reoffending. ROC analysis computes an AUC (area under the curve) by plotting the sensitivity of a tool against its specificity. The resulting AUC score can range from 0 to 1, ≤0.5 Indicates that the tool is not able to predict any better than chance.

DISCUSSION

There were no studies identified that sought to validate tools designed to assist professionals distinguish between behaviours that are within normal behaviour parameters for a child's development and should not be classified as HSB and those that should be categorised as HSB and for which appropriate assessment and treatment should be sought. This is a notable gap in the evidence base as these tools are important in preventing the potential harm that may result if a child is inappropriately referred and assessed with the consequent concerns regarding stigmatisation. This also raises concerns that professionals do not have validated tools that assist in determining when referral should occur so that behaviours might be addressed and treated at an early stage. There were a number of tools that were identified in the process of sifting the literature and in consultation with content experts. These included the Brook Traffic Light Tool (Yamamoto & Kitan, 2015) and Hackett's continuum model of children and young people's sexual behaviours.

We identified eleven studies (Elkovitch et al 2008, Epperson and Ralston 2009, Griffin et al 2008, Griffin & Vettor 2012, Martinez et al 2007, Prentky 2006, Rajlic & Gretton 2010, Viljoen et al 2007, Viljoen et al 2009, Worling et al 2012, Worling and Langton 2015) that sought to validate the accuracy of the risk assessment tools designed to identify those adolescents who have already committed sexual offences and who would be at risk of reoffending.

There was evidence for the following assessment tools; AIM2, adapted AIM, J-SOAP II, ERASOR, BERS-22, J-SORRAT-II, SAVRY, YLS/CMI, PCL:YV and Static-99. Two further tools were identified (MEGA and ASSET) but no validation data was identified to include them in this review.

The evidence demonstrating accuracy in predicting sexual recidivism was not consistent across studies for any of the tools. Three studies, including two high quality studies (Rajlic and Gretton 2010, Worling et al 2012, Worling and Langon 2015) found that the ERASOR (total score) predicted sexual recidivism significantly better than chance but one study did not significantly predict future sexual reoffending. J-SOAP-II was also found to accurately predict sexual recidivism, however this was supported by only two poorer quality studies (Martinez et al 2007, Prentky 2006). AIM2 demonstrated effectiveness in predicting sexual recidivism in adolescent males of normal intellectual ability and in adolescents males with impaired intellectual ability.

The adapted AIM was specifically designed for use with adolescent males with impaired intellectual ability, and was also able to predict adolescent sexual re-offenders with significant greater accuracy than chance. It appeared to have no additional advantages over AIM2 in predicting sexual and non-sexual re-offending in adolescent males with impaired intellectual ability.

Both the J-SOAP-II and ERASOR were also found not to be effective in predicting sexual recidivism in one study (Viljoen et al 2007). The effectiveness of the tools to predict nonsexual recidivism was also contradictory. There was a greater body of evidence for the J-SOAP-II and ERASOR tools which were both shown to be able to predict future nonsexual recidivism. However, one study (Viljoen et al 2007) did not replicate these findings. The BERS-2 tool is used to identify personal strengths and focus on protective factors. The ability of this tool, to predict factors that protect from future reoffending was evaluated by Worling and Langton (2015). The BERS-2 tool consists of five scales, one of which (Affective Strength – measuring capacity for emotional intimacy) that when highly scored was found to be a factor predictive of protecting from future sexual re-offending and one (School Functioning – measuring aspects of competence with school) in which higher levels were predictive of protecting from future non-sexual re-offending. However, Worling and Langton (2015) conclude that BERS-2 did significantly enhance the predictive accuracy of the ERASOR for sexual re-offending.

The J-SORRAT-II tool was also shown to predict sexual recidivism in one low quality study (Epperson & Ralston 2009). None of the remaining tools evaluated were able to accurately predict future sexual recidivism in adolescent sex offenders who have participated in a treatment programme.

The SAVRY tool was able to predict serious nonsexual violent offences, but more reliable in predicting serious nonsexual violence in older youth (Viljoen et al 2007). It may be that tools designed for adolescents aged 12 to 18 do not function equally well for younger and older adolescents. For example, young adolescents may receive higher scores on items on these tools (e.g. impulsivity, lack of empathy) because of their developmental stage rather than more stable characteristics that are indicative of re-offence risk.

There were many mixed findings, and one factor that may contribute to these mixed findings is the fact that rates of sexual reoffending tend to be low, which may make sexual reoffending challenging to predict.

Research with sexually abusive adolescents often appears to focus on the prediction of sexual offence recidivism. However, given that sexually abusive adolescents are three times more likely to commit general rather than sexual reoffences (McCann and Lussier 2008), it appears equally important to examine which tools can predict broader forms of reoffending in this population. It is not clear whether a single assessment tool can predict sexual and general reoffending among sexually abusive youth or if a combination of sex offence-specific tools (such as the ERASOR) and more general tools (such as the YLS/CMI) are needed.

It is also important to examine the characteristics of individuals with whom the tools are more or less effective, i.e. whether certain youth characteristics moderate the predictive validity of a tool. For example, false positives may be more common among young adolescents who may receive inflated scores on certain items (e.g. impulsivity) because of their young age and developmental immaturity rather than stable characteristics associated with long-term risk. All of the HSB assessment models identified have been developed for use with young men. There is no evidence to support their use or their validity for assessing the risks and needs of young women with HSB.

Similarly, it is important to highlight the lack of any formal assessment tools for pre-adolescent children with harmful sexual behaviours and the potential dangers of using models based on research with adolescents on this group, again due to the significant developmental differences between pre and post pubescent children. Chaffin and colleagues (2002) suggest that it is more important to assess younger children's environment, in order to gain a picture of the child's overall social ecology, than it is to focus on the range of offence related variables typically used in juvenile assessment models.

There is a lack of evidence assessing the appropriateness and effectiveness of assessment tools in children with learning disabilities, for girls and adolescent females, and for those in ethnic minority groups. It is also very limited in terms of evaluating their effectiveness in promoting multi-agency working and appropriate treatment referrals.

The growing use of the internet as a means of exposing children and young people to risk of harm, but also as a vehicle for HSB should be highlighted. The ability of the assessment tools to adapt to these changing patterns of behaviour was not described.

Conclusion

The current review has identified 21 risk assessment tools used in residential and community settings. These instruments range from those tools widely used across the UK, such as the AIM2 initial assessment model, and across North America (such as J-SOAP II or ERASOR), to practice tools often developed by individual specialist services as part of their operational practice. The predictive validity of the range of tools to assess levels of risk of future HSB remains uncertain, not least because the tools vary in the specific factors they seek to measure. This is further compounded by the fact that many more recently proposed tools do not merely seek to predict risk of sexual recidivism, but also seek to measure the likelihood that an individual will reoffend in a non-sexual way. Additionally, there has been a shift from tools seeking primarily to assess risk of reoffending, to a more holistic approach to assessment practice which includes not only risk assessment, but also the assessment of strengths and needs. This is consistent with the developments in practice over the past decade which have seen a move from a deficits approach which identified abnormal and problematic elements of a young person's presentation, to more holistic models which emphasise strengths and competencies alongside problems and deficits.

Bonner et al (1998) and Morenz and Becker (1995) argue that when assessing the needs and risk of youth sexual offenders it is important to incorporate comprehensive evaluation including clinical interviews with the young people and their carers, psychological and psychometric tests and questionnaires. A holistic approach to assessment helps professionals to focus on areas which might be related to the risk of sexual offending including attitudes, values, morale, social skills, psychological functioning and sexual knowledge (Becker and Kaplan, 1993). In addition to this, professionals need to consider data from victim statements, court records and mental health reports. In view of the heterogeneous nature of children and young people with harmful sexual behaviour, the review focuses on the investigation of various assessment tools and measures aiming to comprehensively assess individuals. Such tools tend to include assessment of a young person's needs (psychological, social, cognitive, and medical), family relationships, risk factors and risk management possibilities.

Recommendations for practice and research:

Practitioners need better evidenced models to support both the identification of problematic and harmful sexual behaviours (and their distinction from normal sexual behaviours) and also to assess the risk that a child or young demonstrating HSB will re-offend.

The Traffic Light Tool is well regarded by practitioners as a tool which assists in the overall assessment of the level of concern that should be raised in respect of a child or young person's sexual behaviours. At present, it is based on clinical consensus and practice based wisdom. Research should test the accuracy of the model empirically amongst clinical and non clinical samples of children and young people.

Currently, interagency practice across youth justice and social care settings with young people with HSB in the UK is dominated by the use of the AIM2 assessment tool. The originators of AIM2 are clear that it is an initial assessment model which is meant to guide professional decision making. The tool was developed on the basis of evidence from empirical studies of recidivism and also on practice wisdom. Two studies of moderate quality conducted in the UK (Griffin et al 2008, Griffin & Vettor 2012) have sought to evaluate the effectiveness of AIM2 and the adapted AIM to predict risk of sexual recidivism in adolescent male sex offenders. The evidence suggests that it was a reliable tool in predicting sexual recidivism in adolescent males with and without intellectual disabilities. Although this evidence is limited, and further research is warranted, its development in the UK makes in a particularly valuable tool for the UK context.

Internationally, the two tools with the highest degree of empirical support are ERASOR and J-SOAP II, though the evidence demonstrating the accuracy of both tools for predicting sexual recidivism is not consistent across studies. Further studies are needed on larger samples to gain a more consistent view of the utility of these models. At present, no studies have compared the use of different models with the same samples in order to test their relative value in predicting both sexual and non sexual recidivism and in informing need. Such research would be valuable for the field.

In the absence of a more consistent evidence picture, the best approach currently may be for practitioners to use AIM2 or to use both ERASOR and J-SOAP II concurrently. In each case the developers of the tools recommend that practitioners should use them alongside their own clinical judgement in order to inform a perspective on risk and need. Therefore, a combination of both of the tools would appear to be the best way of checking and informing clinical judgement.

A qualitative evidence synthesis of attitudes, barriers and facilitators to assessment of children and young people who display harmful sexual behaviour.

AIMS AND BACKGROUND

Objectives and Rationale

Children and young people who harm others may pose a future risk to children other than their present victim. The safety of their victim and other children is therefore paramount. However children and young people who harm others may have themselves suffered considerable disruption in their lives, been exposed to violence within the family, witnessed or been subject to physical abuse or sexual abuse, have problems in their educational development and may have committed other offences.

Those who work with children and young people who harm others recognise that these children are likely to have considerable needs themselves. Early and effective assessment of children and young people who sexually harm others may play an important role in protecting children and in opening up access to a therapeutic pathway and appropriate service provision.

As reported by Miccio-Fonseca & Rasmussen (2011) clinical interviews have no better than chance in predicting whether a sex offender will reoffend. The authors highlight a consequent need for standardized tools which must be individualized to age, gender, and intellectual capacities. Actuarial risk assessment tools give probabilistic estimates of reoffending over time (e.g., 5 years) (Miccio-Fonseca & Rasmussen, 2011). One actuarial tool for adolescents exists: *Juvenile Sexual Offence Recidivism Risk Assessment Tool-II* (Epperson, Ralston, Fowers, DeWitt, & Gore, 2006). Empirically grounded risk assessment tools are based on risk factors found significantly associated with risk for inappropriate and/or sexually abusive behaviour across studies (Miccio-Fonseca & Rasmussen, 2011). Validated risk assessment tools include: (a) *Juvenile Sex Offender Assessment Protocol (J-SOAP-II)*, (Prentky, Harris, Frizzell, & Righthand, 2000; Prentky & Righthand, 2003); (b) *Estimate of Risk of Adolescent Sexual Offence Recidivism (ERASOR, Version 2.0)*, (Worling & Curwen, 2001; Worling, 2004); and (c) *Multiplex Empirically Guided Inventory of Ecological Aggregates for Assessing Sexually Abusive Adolescents and Children (Ages 19 and Under) (MEGA^r)* (Miccio-Fonseca, 2006, 2009, 2010). Few specialist tools are designed specifically for children or young people engaging in sexually harmful

behaviour, and only one of these, the AIM assessment framework and its successor AIM2, has been validated in the UK. ASSET is another UK-based more general tool that has been separately validated and is incorporated within the overall AIM2 assessment approach. For assessment tools to be effective requires that they are perceived as useful by practitioners, by the adolescents themselves and by their carers. They must be acceptable and be delivered within an integrated pathway of care. Qualitative evidence on the assessment process and the perceived value of specific tools is elusive and diffuse. There is a need to review and synthesise lessons learned from this evidence base to inform delivery of the assessment process and development of appropriate tools.

This qualitative evidence synthesis (qualitative systematic review) seeks to complement an effectiveness review by examining existing published and unpublished qualitative research to establish which methods or components of assessment are viewed as acceptable or useful by children or adolescents who display harmful sexual behaviour, their parents or carers, health or social care professionals and health or social care managers and what considerations should be addressed when seeking to implement such assessments.

Review Questions

The overall review question, quantitative and qualitative evidence combined was:

What types of assessment are effective and acceptable for children and young people who display harmful sexual behaviour (HSB)?

Within this overall question the qualitative review component sought to identify data on the assessment process from diverse stakeholder perspectives (i.e. young people, their family and carers, health and social care professionals and service managers). This would include, but not be limited to:

- Barriers and facilitators to assessment
- Issues relating to feasibility and implementation
- Issues relating to cost implications were not included in the qualitative evidence synthesis in recognition of the separate economic analysis being conducted as part of the NICE guidance programme.

METHODS

Identification of evidence

Searches have been conducted across a range of multi-disciplinary bibliographic databases (See below). Search terms were developed from the scoping search and in discussion with the NICE team. Thesaurus and free-text terms were utilised, relating to the population (children and young people who demonstrate harmful sexual behaviour) combined with terms relating to assessment. Assessment terms were not specifically required for implementation of the qualitative research set. The presence of data on assessment was established at a subsequent stage of title and abstract screening, once the retrieved item was positively identified as qualitative research or a survey containing qualitative data. All searches were limited to English Language, Humans, and the publication time span of 1990-present.

Databases searched

The databases listed in Table 7 were searched in August 2015 for evidence to provide insights on the assessment process in general or on specific tools or methods of assessment.

Table 7 - List of database sources searched for the combined quantitative and qualitative reviews

MEDLINE via Ovid 1946-March Week 4 2015
Ovid MEDLINE In-Process & Other Non-Indexed Citations March 26, 2015
Embase via Ovid 1974 to 2015 March 26
Cochrane Database of Systematic Reviews via The Cochrane Library: Issue 3 of 12, March 2015
Database of Abstracts of Reviews of Effect via The Cochrane Library: Issue 1 of 4, January 2015
Cochrane Central Register of Controlled Trials via The Cochrane Library: Issue 2 of 12, February 2015
Health Technology Assessment Database via The Cochrane Library : Issue 1 of 4, January 2015
NHS Economic Evaluation Database via The Cochrane Library: Issue 1 of 4, January 2015
Science Citation Index Expanded (SCI-EXPANDED) --1900-present and Social Sciences Citation Index (SSCI) --1956-present via Web of Science
Social Care Online 1980-March 2015
PsycINFO via Ovid 1806 to March Week 4 2015

Social Policy and Practice via OvidSP 201503

EPPI-Centre - Bibliomap (mostly pre-2011), Dopher (2006-March 2015), TRoPHI (2004-March 2015)

The Campbell Library 2004-2015 (Volume 11)

Inclusion of relevant evidence

For inclusion in the qualitative evidence synthesis a paper either had to (1a) represent a qualitative research study, using accepted methods of qualitative data collection and analysis or (1b) represent a survey seeking to elicit views on qualitative aspects of the assessment process. Studies should either (2a) directly examine the experiences of adolescents, parents or carers, health or social care professionals or managers relating to the overall assessment process for adolescents with harmful sexual behaviour or (2b) examine experiences of adolescents with harmful sexual behaviour, parents or carers, health or social care professionals or managers relating to use of a specific tool or instrument. In this way data could inform an understanding of either specific assessment provision or of the experience of assessment more generally.

Methods of analysis/synthesis

To enable data to be processed in an efficient manner the team identified four categories of assessment tools. Data were categorised according to the following types of assessment:

- 1) General assessment tools specific to young people with HSB (i.e. to help with the identification of behaviours) e.g. Brook traffic light tool, Hackett continuum model, Ryan red/ green flag model, Finkelhor's 4 pre-conditions model
- 2) Risk assessment tools specific to young people with HSB (i.e. to estimate risk of recidivism) e.g. J-SOAP II (Prentky & Righthand, 2003), AIM2 (Griffin et al, 2008), ERASOR (Worling & Curwen, 2001), JSORRAT (Epperson et al, 2005)
- 3) Generic assessment tools for children and youth at risk of harm or harming others (including but not specific to, young people with HSB) e.g. DH Assessment Framework (Department of Health, 2000), ASSET (Youth Justice Board), SAVRY (Borum et al, 2002)
- 4) Component tools (i.e. tools not specific to HSB but which address factors implicit in HSB and which might form part of an overall assessment) e.g. Beck Depression Inventory, PCL-YV, Trauma Symptom Checklist, Multiphasic Sex Inventory, Beckett's ASAP measures, etc.

These four specific categories were prefaced by a more substantive general category populated by qualitative data relating to the assessment process itself where specific named tools were not directly referenced.

Quality assessment

Quality Assessment was conducted in accordance with the current version of the NICE manual procedures for assessment of qualitative studies. All questions were coded in a Google Form which was completed during data extraction. The combined assessment of each study was then used to inform the allocation of overall study quality, indicated using the agreed ++, + and – notation.

Data extraction

Data was initially extracted against a generic data extraction form, handled via Google Forms. Data was exported to an Excel spreadsheet to facilitate manipulation of the data and identification of patterns to inform the synthesis.

Data analysis and synthesis

Thematic synthesis was used to analyse data on general aspects of the assessment process (Section 5.2). The resultant themes were then examined within the accounts of specific tools (ASSET – Section 5.4) and approaches (AIM/AIM2 – Section 5.3). Additional themes were identified from the evaluations of these specific approaches and described using narrative approaches.

SUMMARY OF INCLUDED STUDIES

Identified studies

Two relevant papers were identified from searching of the electronic databases (Griffin et al, 2008) and subsequent follow up of references (Griffin & Beech, 2004). This low hit rate from a formal search procedure indicates that the term “assessment” is problematic, carrying multiple meanings (false positives). In addition data relating to assessment was only identified from examination of full-text reports with the concept of assessment not being clear from either title or abstract (false negatives). Four studies (Belton et al, 2014; Deacon, 2015; Geary, 2007; Hall, 2006) were identified from the qualitative evidence synthesis of interventions (Campbell et al 2016) commissioned at the same time as this review. Citation searching of included studies on Google Scholar and searches of a “sensitive database” of items retrieved using additional non-core terms revealed a further five qualitative studies (Baker et al, 2005; Geary et al, 2011; Hall, 2010; Ladwa-Thomas & Sanders, 1999; Roberts et al, 2001). This means that 11 papers are included in the qualitative evidence synthesis. (see flow diagram , figure 1, for a description showing the process of identifying relevant studies). Thirty eight papers were excluded as abstracts containing insufficient detail of qualitative data, dissertations or other items that were unavailable or items that, on close inspection of the full text, were not eligible for inclusion (See list of excluded studies).

Included studies

Study characteristics

Included papers covered the period from 1999-2015. The 11 included papers report initiatives from only the two following countries, presented in order of frequency:

United Kingdom [Nine reports]: Baker et al, (2005); Belton et al (2014); Deacon (2015); Griffin & Beech (2004); Griffin et al (2008); Hall (2006); Hall (2010); Ladwa-Thomas & Sanders (1999); Roberts et al, (2001)

New Zealand [Two reports]: Geary (2007); Geary et al (2011)

Most papers sampled from health and/or social care professionals (7). Studies including consumer perspectives included both the perspectives of children/adolescents (4 studies) and of parents/carers (2 studies). Two of these papers also included the perspectives of health/social care professionals. As might be expected, given the operational focus of the review question, few papers (n=1) sampled from managers. However some papers did report the

implications of the assessment process in terms of managerial support etcetera. The distribution of perspectives by study is indicated in Table 8. The absence of papers from the United States is surprising and there is an attempt to explain this in the Discussion section.

Table 8 - Perspectives captured in Included Studies

Study Identifier	Children/ Adolescents	Parents/Carers	Health or Social Care Professionals	Managers
Baker et al (2005)			✓	✓
Belton et al (2014)	✓	✓		
Deacon (2015)			✓	
Geary (2007)	✓	✓	✓	✓
Geary et al (2011)	✓	✓		
Griffin & Beech (2004)	✓	✓	✓	
Griffin et al (2008)			✓	
Hall (2006)			✓	
Hall (2010)			✓	
Ladwa-Thomas & Sanders (1999)			✓	
Roberts et al (2001)			✓	✓
	4	4	9	3

Eight of the 11 included studies used semi-structured interviews (Table 9). Five studies used questionnaires/surveys. Three of these studies (Baker et al, 2005; Hall, 2010; Roberts et al, 2001) used a questionnaire followed up by more detailed interviews. Two studies used additional methods but not specifically as a source for the data on assessment.

Table 9 - Data Collection Methods used in Included Studies

Study Identifier	Semi-Structured Interviews	Questionnaires/ Surveys	Observation (Other)
Baker et al (2005)	✓	✓	Review of assessment reports
Belton et al (2014)	✓		
Deacon (2015)	✓		✓ (i.e. Recordings)
Geary (2007)	✓		
Geary et al (2011)	✓		
Griffin & Beech (2004)	✓		
Griffin et al (2008)		✓	
Hall (2006)	✓		✓ (Review of case files)
Hall (2010)	✓	✓	
Ladwa-Thomas & Sanders (1999)	✓		
Roberts et al (2001)	✓	✓	
	10	3	2

The majority of studies (with only one exception) used a single method of data collection so triangulation across methods was not possible. The details of the methodology and populations of the included studies are summarised in Table 10. Full study details are presented in the evidence tables (Appendix 4).

Given the UK predominance in included studies it is not surprising to see that four of the nine studies reported within the context of the AIM or AIM2 frameworks (Belton et al, 2014; Deacon, 2015; Griffin & Beech, 2004; Griffin et al, 2008) with a further two reports evaluating the ASSET tool (Baker et al, 2005; Roberts, 2001). Other studies mentioned additional instruments but mainly in passing (i.e. not as a focus for research or evaluation). These were ERASOR and J-SOAP (both Geary 2007); ASSET (Hall, 2010) and the DH Framework (Hall, 2006; Ladwa-Thomas & Sanders, 1999). The study by Geary et al (2011) did not reference a specific instrument but might be interpreted as covering the same population as the more detailed earlier study (Geary, 2007). As a consequence of the above there is more qualitative data relating to generic aspects of assessment than, with the exception of AIM/AIM2 for individual tools, instruments or approaches. A particular omission, therefore, is qualitative data relating to

some of the quantitative common tools used to predict recidivism, the focus of much research and evaluation in the United States.

Table 10 - Populations, Aims and Settings of Included Studies

Study Identifier	Aim	Method and population	Location	Assessment Methods/Tools
Baker et al (2005)	To provide a detailed account of practitioners' views on the purpose, design and use of ASSET	Social work practitioners and operational managers.	United Kingdom	ASSET
Belton et al (2014)	To understand how manualised programme for males aged 12-18 with harmful sexual behaviour (HSB) is used and experienced in a social care context.	Young people and their parents or carers	United Kingdom	AIM/AIM2
Deacon (2015)	To understand how CSCS deal with referrals of children with SHB; Reflections of social work (SW) practitioners when working with these families; User (parent/carer) views about how cases were managed; Parent/carer experience of SW interventions. Best practice recommendations to inform effective intervention by SW practitioners, and training to be offered	Social work practitioners	United Kingdom	AIM/AIM2
Geary (2007)	To examine the New Zealand variation of Adolescent Sexual Offender treatment by providing detailed information about the characteristics, operation, and services provided by the local programmes	Social work professionals, adolescents and parents	New Zealand	ERASOR; J-SOAP
Geary et al (2011)	To identify consumer perspectives of strengths and weaknesses of programme delivery at three community programmes for sexually abusive youth	Adolescents plus a range of caregiver roles (parent, extended family member, step-parent, placement caregiver)	New Zealand	Not specified

Study Identifier	Aim	Method and population	Location	Assessment Methods/Tools
Griffin & Beech (2004)	To evaluate effectiveness of AIM framework: to assess use and effectiveness of multi-agency approach promoted through AIM framework; to assess usefulness of AIM model as a screening tool to assist practitioners; to evaluate level of accuracy of strength and concern continuums and outcome matrix; to identify how 10-step framework is used in practice and how it impacts on professionals; to look at how the AIM assessment impacts on young people and their families/carers.	Professionals, young people and families/carers	United Kingdom	AIM
Griffin et al (2008)	To describe the AIM2 assessment framework	Professionals	United Kingdom	AIM2
Hall (2006)	To see how one social services department had responded to national guidance, issued in Working Together (DoH, 1991),	Social Workers	United Kingdom	DH Framework
Hall (2010)	To test theories on meanings of working holistically with young people who have sexually harmed; To identify further meanings, benefits, challenges and implications of working holistically with this client group; and To produce recommendations for policy and practice relating to holistic working with young people who have sexually harmed.	Staff from the Youth Offending Team and associated agencies.	United Kingdom	ASSET
Ladwa-Thomas & Sanders (1999)	To explore social worker definitions of abusive behaviour, views as to the causes of young people abusing others, social work intervention and personal resources needed to work with young abusers.	Social workers	United Kingdom	DH Framework
Roberts et al (2001)	To provide a detailed account of practitioners' views on the purpose, design and use of ASSET	Social work practitioners and operational managers.	United Kingdom	ASSET

Study methodology and quality appraisal

The results of quality assessment for the eleven studies are presented in Table 12. Three papers were rated high (++), six moderate (+) and two low (-) (see Table 11). Areas where papers received low ratings include: the unclear role of the researcher; the thin description of context; the uncertain reliability of analysis; and the lack of 'richness' of the data reported. Of the nine UK studies one was judged of high quality (Griffin et al, 2008); four were judged as moderate quality (Baker et al, 2005; Belton et al, 2014; Deacon, 2015; Roberts et al, 2001) and the remaining two were assessed as low quality (Griffin & Beech, 2004; Ladwa-Thomas & Sanders, 1999). The moderate overall quality of the UK studies and their contribution to the resultant synthesis indicates that the body of evidence possesses high relevance but moderate rigour (see section 4.4 below).

Table 11 - Overview of the study quality of the included qualitative studies.

Study design	N identified	Quality Rating		
		++	+	-
Qualitative Studies	11	3	6	2

Table 12 - Quality Assessments for Included Qualitative Studies

Reference	Qualitative approach	Data Collection	Study Purpose	Study Design	Role of Researcher	Context	Reliable Methods	Rigorous Data Analysis	Rich Data	Reliable Analysis	Convincing Findings	Relevant Findings	Conclusions	Cl Co Re
Baker et al (2005)	Appropriate	Appropriate	Clear	Defensible	Unclear	Not Sure	Reliable	Rigorous	Rich	Not Sure	Convincing	Relevant	Adequate	Ap
Belton et al (2014)	Appropriate	Appropriate	Clear	Defensible	Unclear	Not Sure	Reliable	Rigorous	Rich	Not Sure	Convincing	Relevant	Adequate	Ap
Deacon (2015)	Appropriate	Appropriate	Clear	Defensible	Partially Clear	Clear	Reliable	Rigorous	Rich	Not Sure	Convincing	Relevant	Adequate	Ap
Geary (2007)	Appropriate	Appropriate	Mixed	Defensible	Unclear	Clear	Reliable	Rigorous	Rich	Not Sure	Convincing	Relevant	Adequate	Ap
Geary et al (2011)	Appropriate	Appropriate	Clear	Defensible	Partially Clear	Clear	Reliable	Rigorous	Poor	Reliable	Convincing	Relevant	Adequate	Ap
Griffin & Beech (2004)	Appropriate	Appropriate	Clear	Defensible	Unclear	Not Sure	Reliable	Not Sure	Poor	Not Sure	Convincing	Relevant	Adequate	Ap
Griffin et al (2008)	Appropriate	Appropriate	Clear	Defensible	Unclear	Not Sure	Reliable	Rigorous	Poor	Reliable	Convincing	Relevant	Adequate	Ap
Hall (2006)	Appropriate	Appropriate	Mixed	Defensible	Partially Clear	Clear	Reliable	Not Sure	Rich	Not Sure	Convincing	Relevant	Adequate	N
Hall (2010)	Appropriate	Appropriate	Mixed	Defensible	Unclear	Clear	Reliable	Rigorous	Rich	Not Sure	Not Sure	Relevant	Adequate	Ap
Ladwa-Thomas & Sanders (1999)	Appropriate	Appropriate	Mixed	Not Sure	Unclear	Not Sure	Not Sure	Not Sure	Poor	Not Sure	Not Sure	Relevant	Adequate	N
Roberts et al (2001)	Appropriate	Appropriate	Clear	Defensible	Unclear	Not Sure	Reliable	Rigorous	Rich	Not Sure	Convincing	Relevant	Adequate	Ap

Applicability

Of the 11 included studies nine were conducted in the UK. Five of the UK studies examine an overall approach to assessment, typically performed by social work professionals or Youth Offending Teams, while the four remaining UK studies describe the performance of specific tools (AIM/AIM2 and ASSET). The six richest studies, in terms of data on assessment, are the earlier New Zealand study (the thesis by Geary, 2007), the UK thesis (Deacon, 2015) and the evaluations of AIM/AIM2 (Griffin & Beech, 2004; Griffin et al, 2008) and ASSET (Roberts et al, 2001; Baker et al, 2005).

Of particular note is the fact that by far the large majority of studies has been published within the last decade. This fact, plus the predominance of UK studies, increases our confidence in the applicability of the qualitative findings. Nevertheless the absence of studies examining stakeholder perspectives of the validated instruments that are used for actuarial purposes in the United States to predict recidivism is a limitation of this review.

STUDY FINDINGS

We sought to categorise review findings under five categories:

The Assessment Process

7 of the 9 studies (Belton et al, 2014; Deacon, 2015; Geary, 2007; Geary et al, 2011; Hall, 2006; Hall, 2010; Ladwa-Thomas & Sanders, 1999) referred to aspects of the generic assessment process. In this section the review looks at evidence regarding the core general safeguarding assessment and also the more specific assessment of the HSB itself. ASSETT is a tool commonly used by youth offending teams in the UK for the general assessment of young people with offending behaviour and is not specific to the assessment of HSB. We have included evidence regarding its use in this review. AIM2 is a tool also commonly used in the UK and can be used by all professionals involved in the assessment of HSB.

Assessment process as a means of facilitating multi-agency working

For some practitioners the assessment process is seen as a welcome opportunity to talk to other professionals and thus to gather different types of information relating to the case (Belton et al, 2014):

“...being able to talk to the NSPCC and the practitioners was really helpful and they were very open with me around what I should be looking for, what work could be done, what the young person’s perspective was on things, and dad’s as well, and it helped me to focus on my assessment, which was incredibly useful.”[Social Worker]

By way of contrast, a multi-agency collaborative assessment may simply serve as a tangible reminder of the need for greater co-operation with “more interagency meetings and collaboration with other agencies involved in the care of the young person” (Geary, 2007):

“We could improve assessment feedback by widening the net and having more interagency meetings – where it happens it works really well. It’s a question of time. Sometimes the interagency communicating is done by phone or email. (programme manager)”.

Deacon (2015) reports that it was unusual for social workers completing general safeguarding core assessments to include an analysis of sexually harmful behaviour. One social worker commented that “We didn’t consider it to be necessary to look deeply into the sexualised behaviour in our assessment- that was being handled by the YOT”. A significant proportion of

initial and general safeguarding core assessments were based on limited information; some were superficial and lacked analysis of safeguarding needs or potential risks. As a consequence, needs were underestimated and minimal interventions were provided.

Suggestions for improvement included a post-assessment case conference of all the professionals involved, formalised as part of the assessment. It was also suggested that it would be helpful “if the professionals sorted out business first and then involved the family...It should be professionals only meeting first” (Geary, 2007).

The AIM framework approach (Griffin & Beech, 2004) is conceived as a multi-agency approach. Griffin et al (2008) highlight how inter-agency co-operation has increased through use of a standard assessment framework:

“Through these assessments, good relationships have been built between YOT and social services departments” (YOT practitioner).

There is some evidence that those using the ASSET tool were unwilling to share assessments across agencies (Baker et al, 2005). Concerns related to data protection issues, that other organisations may not understand the purpose of ASSET or might use assessments inappropriately. Examples were also given of organisations being unwilling to accept referrals on the basis of ASSET assessments (Baker et al, 2005). Some or all of these issues may reflect generic concerns for the assessment process. On the other hand provision of electronic versions of the ASSET tool was perceived to increase sharing and to increase the likelihood of repeated assessments over time (Baker et al, 2005). Again these issues may also pertain to a generic assessment process.

Providing better information pre-assessment

Geary et al (2011) identify a need to provide better information for the benefit of adolescents and their parents before the assessment. This should include information about assessment procedures. The location of the assessment is considered important. Many parents and caregivers express a preference for assessments to take place in their own homes (Geary et al, 2011), particularly important for certain ethnic groups. Assessment is seen as the first opportunity to put the adolescent at ease and thus reducing their anxiety both for the assessment and for the programme as a whole. While this initial anxiety or discomfort can ultimately be overcome by the friendly interactions of the staff it is clear that more can be done to facilitate this process. For those professionals undertaking the assessment of HSB, they

require the necessary skills in order build a relationship of trust, engagement and acceptance with the young person.

Developing rapport between practitioner and adolescent

Several reports emphasise the benefits to be accrued from working with a young person on a one-to-one basis. Belton and colleagues (2014) describe how "quieter young people sometimes found it difficult working with two practitioners and preferred one-to-one work". An important consideration here is continuity between the practitioner and the young person. Young people felt that, having "got to know" the practitioner who conducted the assessment, it was important to maintain this dynamic into the therapeutic process. For the practitioner benefits from continuity related to building up knowledge and background about the case. For the young person continuity obviated the need to rebuild a relationship of trust and repeat things they had discussed in the assessment (+).

The intervention review identified the importance of building up trust, rapport and communication between the professional and the young person. The assessment process is seen very much by practitioners as a first opportunity to achieve this:

"The core relation of the assessment is you get to know them, and this is the problem with why the core assessment takes so long. To get as honest a picture as you ever can, you've got to build a rapport with them to get them to the point to talk about the offence..." (Belton et al, 2014)

Geary (2007) reports that "Once the assessment process got underway, getting to know the therapist played a central role in helping adolescents to open up". In many cases adolescents may not find it easy to form relationships with people when they first meet them.

Communicating with the adolescents and the parents at the same level helps to "open up channels of communication" (Geary, 2007). This may involve explaining the questions or rephrasing them in a way that both adolescents and parents understand.

We can conclude that the assessment process contributes to a good working relationship between adolescents and social workers and between parents and social workers. However conducting an initial assessment is *not* a pre-requisite for good relationships; skilled professionals show themselves able to develop good working relationships in a variety of procedural contexts.

Characteristics of assessment reports

Geary (2007) reports that external agency staff appreciated reports that were “professional, prompt, and offered recommendations that were possible and well researched”. It was felt important that these offer a “comprehensive history of the person – the behaviour is seen in context”. One social worker recounted that sometimes they received more information in the assessment report than they had previously been aware of and remarked that “people often felt more comfortable sharing knowledge with the programme than with us”. Increasingly practitioners expressed a need, reflected in the general social work assessment literature, for the assessment process to be able to take place electronically.

Frustrations from assessment reports (Geary 2007) relate to a need for standardisation across programmes:

“I would like to see standardised assessment measures with other agencies as we have kids that go to (other programmes). We’re a small sector – it would create broader understanding. We don’t want to lose our ability to share information and look at other issues”.

In a New Zealand context a need was expressed to incorporate a cultural perspective when conducting assessments (Geary et al, 2007; Geary et al, 2011). While this was expressed specific to the sizeable Maori community this may have similar implications for other ethnic groups. Experience from the AIM2 programme also emphasised the need for cultural sensitivity, particularly when asking potentially obtrusive questions. Sensitivity to the sexual content of questions, which may be too explicit for particular communities, is therefore one such issue to be taken into account (Geary et al, 2007; Geary et al, 2011).. However a further consideration may relate to how the assessment is documented e.g. whether it is acceptable to be taking notes or whether it is preferable to enter material into the computer immediately after an interview. Concerns about language and literacy should also be addressed: “Filling out forms is a reminder of school failure”.

Timing and thoroughness of assessments

Several reports highlight the importance of undertaking an “earlier and more thorough assessment” (Ladwa-Thomas & Sanders, 1999). However it was recognised that this would have time and resource implications (Ladwa-Thomas & Sanders, 1999).

Waiting lists were seen to exacerbate the problem regarding suboptimal delayed assessment...They talked about the difficulties experienced when clients had to be placed on waiting lists some of which could last as long as three to four months (Geary, 2007).

The assessment process was depicted as a type of bottleneck preventing young people from gaining access to appropriate care (Geary, 2007):

The length of time it takes between the referral being made and the acceptance and assessment taking place is not adequate. There are considerable delays... (Residential homes) won't accept until assessments are completed...We can't always place people in safe environments while they are waiting. Even 3 weeks is too long for us ... the school won't let a person back to school until he's accepted.

Frequently assessment is portrayed as a gateway to effective treatment with the strong implication that inadequate or delayed assessment results in inadequate or inappropriate care. . For example Deacon (2015) reports that:

Adoptive Mother was the only one who was really negative about the therapeutic support as she was not happy that a proper assessment had not been completed of her son and she therefore questioned how the therapeutic worker was able to address his needs when this has not been done.

However it should not be inferred that assessment is a static one-off process because this can result in a lack of professional ownership of the results of the assessment by subsequent professionals involved in handling the case (Deacon, 2015).

Team Manager [felt that] the way [cases] were represented to her depended on the value-base of the worker who was handing over the case....and she felt that it was important to question the interpretations given.

I feel my role is to question [the worker] to ensure they are appropriately assessing the situation to ensure the social worker can justify their decisions and not just take the anxiety from other professionals

Typically the issue of delay is framed as an issue of "safety", for both offender and victim(s), perhaps as a blame avoidance strategy or as a stratagem for securing earlier assessment (Deacon, 2015):

It's difficult as the young people aren't safe in their current environment if we don't have placements for them, and.

We have a boy at the moment who needs urgent assessment. The time factor is crucial. He's in a really quite inappropriate environment and he's on the waiting list. It's 12 weeks. We're trying to keep him safe, keep the victims safe and keep the family informed.

In the study by Geary (2007) staff suggested that high risk adolescents should be prioritised for the assessment process. During the assessment process the family requires support, the client needs to be kept engaged and all need to be assisted with interim safety plans (Geary, 2007).

Delayed or omitted assessments are frequently portrayed as missed opportunities for intervention or prevention. For example Deacon (2015) described how:

Many of the cases in this research study, with the benefit of hindsight, indicate strongly that if an assessment had been completed then information is more likely to have come to light that could have prevented the child in question from committing further acts of SHB.

She further argues that use of an appropriate, validated tool would "improve the chances of assessing more accurately whether behaviour is abusive or experimental" and taking this decision beyond the judgement of individual social work practitioners alone. She cites Miccio-Fonseca and Rasmussen (2011) as one example of such research.

Training and support

In the earliest report in this review (Ladwa-Thomas & Sanders, 1999) practitioners felt that "their lack of skills in working with young abusers immobilized them". They felt that comprehensive assessment was "beyond their capabilities, given the lack of skills knowledge and support available to them". While practitioners generally "felt willing to undertake investigation and assessment of abuse, they saw assessment without adequate resources as untenable".

While creating a one-to-one relationship was felt to be important social workers in a study by Hall (2006) cited co-working with another social worker as the most useful factor:

'I think co-working is important . . . because if you challenge an attitude of the young person it would be reinforced by your co-worker.'

'Because again you have someone to bounce ideas off.'

Hall reports that social workers struggled to complete risk assessments:

'I think the difficult bit was actually sitting down and coming up with the information for the conference.'

Sources of support that were valued included informal supervision and external supervision (Hall, 2006):

'Informal and *ad hoc* supervision was informative . . . because the young person was in total denial, so we had to get some ideas about how to approach that mind set.'

Lack of managerial support was also seen as an important issue (Deacon, 2015). One Independent Reviewing Officer related that he could not remember the name of the team manager he had while working on a specific case but:

'what I can absolutely guarantee is that I would have been given the case, and I was a senior practitioner, and I would have went out and managed it and made decisions as I saw. I never felt that there was any input from the managers other than the authorisation of assessments.' (Deacon, 2015)

More recently Deacon (2015) found that training provision was poor:

Independent Reviewing Officer had completed a one day training course prior to him dealing with any cases relating to SHB. However, he expressed concern that social workers in long-term teams are not trained to handle referrals regarding SHB
...Referrals are normally handled by a duty team so they are trained to some extent as to what to do, but social workers in long-term teams are not.

The Children Act (1989) was said to be unhelpful in not identifying adolescents who sexually abuse others as children 'in need'. Some felt that assessment was better undertaken by criminal justice staff, because 'society had already indicated needs for sanctions when individuals offend norms' (Ladwa-Thomas & Sanders, 1999). This may simply be an attempt to push away the problem to someone else, a reflection of lack of sufficient training, or an inability to be able to appropriately draw on previously learnt skills. Nevertheless, it highlights the underlying sense of lack of necessary ability. This links with the further finding by Ladwa-Thomas & Sanders (1999) that "lack of skills in challenging the denial of abusers and carers" was a major concern for practitioners. This study further reports "how to assess risk of reoffending and, more

generally, what pertinent questions to ask when undertaking a comprehensive assessment” as identified skills gaps.

Integration with intervention

Although the reviews on assessment and therapeutic interventions were conceived separately the qualitative evidence emphasises the need for these processes to be integrated and joined up. For example Belton and colleagues (2014) state, with regard to the manualised treatment programme, that “If the same practitioner who had undertaken the assessment was now also delivering treatment, some of the material in the engagement module was felt to be unnecessary, as the practitioner and young person had already formed a working relationship”. Not only might this save time but it also demonstrates responsiveness to the observation that “It could...feel strange for young people who had ended the assessment by talking in detail about the HSB, to then move back to more general material when they were ready to move the work forward.” There was seen to be a particular need to achieve integration between the assessment in those cases where the NSPCC carries out the assessment, there could be better integration between the assessment and the programme (Belton et al, 2014).

Need for a holistic approach, particularly involving the family:

In a thesis exploring the need for more holistic working with adolescents who have sexually harmed others, Hall (2010) reports that the significance of the family and their influence was raised by many respondents before being mentioned by the researcher: “to work holistically is, it’s I mean the word fullness ... in terms of assessment, but also in terms of delivery ... you’re looking at every aspect of child development and at ...every aspect of that family and the changes and the impact, ... looking with peers and other influences which obviously for young people is key”.

Need to explicitly link assessment to provision:

In the study by Hall (2010) it was considered important to be able to explicitly link the assessment process to provision of services or assistance that addresses a family’ needs, including “appropriate housing, mental health problems, parental relationship difficulties, poverty, parental offending, stigma from the local community, health needs of siblings and other behavioural problems”:

“One YOT worker explained (74): “the complexity of some of our families is massive. ... What tends to happen is that 27 agencies say, are involved with a family... They go along to the house, and they find the parents aren’t in, and so they don’t go back, or they sit

there and listen to the story and it's too overwhelming and they back off. The challenge is to work out how we can actually meet those families' needs, because ... *they're used to being assessed but not being provided for*, I think".

Interestingly, feedback from the ASSET project similarly highlighted the frustrations of practitioners who were able to identify needs but, structurally, were unable to address them.

As with manualised treatment (See Review One) a standardised method of assessment was seen to offer both strengths and weaknesses. Practitioners want a process that is comprehensive and in which they could have confidence. They particularly want to be able to differentiate experimentation from harmful behaviours that might lead to re-offence (Deacon, 2015).

At the same time there is a consistent need expressed in the qualitative evidence to be client-centred showing flexibility and adapting to the individual needs of the client (Deacon, 2015):

Adoptive Mother agreed that social workers should be able to adapt their practice when it was needed as her family already had an adoption social worker and a therapeutic social worker....Her view was that both social workers already involved knew the family well...so why could CSCS not adapt to allow one of them to complete the assessment rather than having another person involved who did not know the family and would have to go through everything again?

However, she felt it can help the assessment for there to just be one social worker as they get the information from both sides which assists them in making an assessment,,,if the alleged victim and alleged perpetrator are from different families then they should 'absolutely' have a separate social worker, but not if they are from the same family.

Tension between Victim and Offender roles

As found in the review of therapeutic interventions (Review 1) ambiguity was felt around the fact that the adolescent was both offender and victim (Deacon (2015):

Independent Reviewing Officer described how he did view Child6 as vulnerable, as during his assessment he 'tried to focus on the wider family dynamics and how they had not put the right boundaries in place, and how the family might be able to support him through that as well as minimising the risks in the future'.

On the other hand Deacon (2015) describes how one senior social Worker expressed concern about social workers viewing children who display SHB as vulnerable victims - not viewing an adolescent as a possible perpetrator of sexual abuse meant they did not address his behaviour earlier, which could possibly have prevented an escalation in his behaviour. This was echoed by an Independent Reviewing Officer:

I know we have the saying “innocent until proven guilty” but as a practising social worker I would say we have received some serious allegations here – is there any reason to disagree with what they’re saying first of all?

It is important to recognise that initiating the process of assessment may result in stigmatisation of the alleged offender and that there are risks associated with assessment. Attempts to safeguard children could place the alleged perpetrator at risk themselves (Deacon, 2015). This was described by one Social Worker as a ‘grey area’ requiring that they consult with the Legal Department on a number of occasions due to its complexity:

“Decisions kept chopping and changing as to whether we should be informing parents, whether we should be opening them up on our system, whether assessments should be undertaken...”

However the corollary is that a failure to conduct a risk assessment can result in action being taken without sufficient justification, for example in an alleged victim having to move placements when risk had not be convincingly established (Deacon, 2015). Furthermore if a risk assessment has not been undertaken then when an offender is released following conviction there is no documentary basis for offering therapeutic support or protective measures for previous victims (Deacon, 2015).

General assessment tools specific to young people with Harmful Sexual Behaviour

No included studies examined the use of general assessment tools for assessing sexual behaviour (e.g. Brook Sexual Behaviours Traffic Light Tool (Brook, 2012), Hackett continuum model (Hackett, 2011). Finkelhor’s 4 pre-conditions model (Finkelhor, 1984) within an adolescent population exhibiting harmful sexual behaviour.

Risk assessment tools specific to young people with HSB (i.e. that estimate risk of recidivism)

One study, represented by two papers (Griffin & Beech, 2004; Griffin et al, 2008), reports an evaluation of the AIM assessment protocol framework. The AIM2 assessment framework and procedures are designed to assist professionals in assessing children who are alleged to have committed a sexual assault or admitted to undertaking sexually harmful behaviour. The framework adopts an explicit multi-agency approach. It incorporates concepts from the DoH 'Framework for the Assessment of Children in Need and their Families' used by Children's Social Care and other agencies, as well as the 'Asset' framework used by YOS (See section 5.4). Its use is intended to fit within the timescales agreed by the criminal justice and child welfare systems. Deacon (2015) describes a context in which the AIM2 assessment is used as a standard assessment and reports rich data on the practicalities of its use.

Training and Support

Deacon (2015) mentions issues around training for completing AIM assessments:

While Social Worker had attended a one-day training course two years prior to this interview she described it as 'not particularly helpful'. She said this was about completing AIM assessments, but she had not experienced any specific training just relating to children who display SHB, and...did not think her Local Authority offered any training in this area.

Participants generally observed a shortage of appropriate training in this area. Training needs became acute when they faced a particularly challenging situation. Reading was one strategy to address this perceived deficiency:

I found that, because of the case I had this year, when I actually looked to see if there was any relevant training, that there was actually nothing if I'm being honest! And I felt that that would have helped a lot with this case. I felt that I had to go away and do a lot of reading myself about this specific area (Deacon, 2015).

Managers found it equally difficult to identify appropriate training. In some cases managers overestimated the numBERS-2 of staff who had received training when compared to data elicited by the project, particularly as AIM training was optional. On the job exposure to challenging situations was seen as equipping professionals for the further demands of their roles (Deacon, 2015). Local authorities simply sought to obtain the minimum required number

of social work staff for training in the expectation that they would then support their colleagues. This was particularly the case given that the AIM assessment was seen as the responsibility of the Youth Offending Teams, not the social care practitioners.

Lack of training in AIM assessment was also expressed as an obstacle to prompt assessment. Even where social workers recognised a need for an AIM assessment they had not been trained to perform it. In one specific case the resultant delay was viewed as critical to a negative outcome for one child (Deacon, 2015)

Perhaps due to the lack of training in AIM assessments for social workers there was “a general sense of confusion from social workers as to the purpose of AIM assessments and how these should be applied in practice” (Deacon, 2015). Informants believed that this was a problem with the existing training offered, which did not really explain the circumstances in which social workers should use them, only how they should be used with adolescents. The term ‘AIM assessments’ (Morrison and Henniker, 2006), though thought to be common knowledge in general social work practice masks the fact that few practitioners seem to know what they actually are or how they should actually be used. Also, Deacon (2015) reports that no other tools were identified in relation to how to deal with younger children who display SHB.

Assessment process as a delay to treatment:

Deacon (2015) reports how the assessment process can act as a source of delay to appropriate intervention:

When asked what other options there are for the assessment of children who display SHB who have not admitted to or been convicted of the offence (as is required before an AIM assessment has been completed) 20_Team Manager’s response was ‘It is difficult because a lot of the agencies who would work with a child who has sexually harmed another child won’t pick it up because there has been no conviction or admission and they’ve said they can’t work with a child who is in denial but I feel that’s wrong...’. She felt it was important that therapeutic intervention began as soon as possible rather than ‘going round in circles’ about children being out of the home or in a settled placement first..

Another Team Manager in Deacon (2015) expressed concerns for a child for whom:

“therapeutic intervention ... had so far been delayed because of the ongoing police investigation and AIM assessment, because he could have a ‘completely different view of the world’ caused by ‘living in that environment’ with his carer”.

Tension between criminal and therapeutic requirements:

The assessment process is seen as a scenario where potential tensions between criminal and therapeutic considerations come to a potentially critical point. Deacon (2015) describes how a Senior Social Worker felt that his remit was to work ‘alongside’ the criminal justice process but that he was unable to complete his work while proceedings were being completed. This informant described this as ‘very frustrating’ but recognised this as necessary because of having to work within the legal context because he ‘wouldn’t want the criminal case to fall apart because of something I’d done’. The informant’s team manager attributed this to a focus not on the “criminal side of the crime or the needs of the victim” but on the needs of the offender himself. She felt this was achieved by completing an AIM assessment. Her comments confirm the uneasy balance to be negotiated when considering the alleged perpetrator’s needs and comparing them to the safeguarding of other children – what has been described as “the ambiguity of juvenile sexual offenders” (Morrison and Henniker, 2006; Harrison, 2009).

Substantial consultation took place in advancing the AIM assessment towards AIM2 (Griffin et al, 2008). Feedback on the draft AIM2 assessment revealed that practitioners considered that the assessment provided a rich platform to guide interventions and treatment planning, especially with regard to dynamic factors (Griffin et al, 2008). Practitioners who completed multiple assessments commented that with practice and experience, the undertaking of a draft AIM2 assessment became easier and took less time (Griffin et al, 2008).

Deacon (2015) describes how some social work practitioners were aware of AIM assessments but lacked understanding as to whether they could usefully be used in practice. A structural barrier was that children actually have to admit to or be convicted of an offence in order for the AIM assessment to happen (AIM website, 2011). Deacon (2015) demonstrated that, for a large number of referrals, children do not admit to the allegation. She points out that where there is evidence that an offence has taken place but this is not substantive enough for the criminal justice system) social work professionals lack alternative tools in order to assess and support these children, potentially creating an assessment limbo.

Generic assessment tools for children and youth at risk of harm or harming others

A limited number of studies (n = 2) examined generic assessment tools and processes. Deacon (2015), having primarily focused on the AIM assessment (see above) makes passing reference to the Department of Health Assessment Framework (2000) which, along with ASSET is incorporated within the AIM2 approach. In addition evaluation of the ASSET framework included a substantive qualitative component.

DH Assessment Framework

One study (Deacon, 2015) reports that generally social work practitioners experience confusion about what to do when confronted with cases concerning children displaying Harmful Sexual behaviour. Specifically she flags that “The generic assessment (Department of Health, 2000) does not specifically recognise the presenting issues of complexities of these particular children”.

ASSET

Asset is a structured assessment tool used by Youth Offending Teams (YOTs) in England and Wales with all young offenders who come into contact with the criminal justice system (Youth Justice Board, 2003; Baker, 2004). It aims to look at the young person’s offence or offences and identify a multitude of factors or circumstances – ranging from lack of educational attainment to mental health problems – which may have contributed to their behaviour. Information gathered from Asset is used to inform court reports so that appropriate intervention programmes can be drawn up. It also highlights particular needs or difficulties the young person has, so that these may be addressed. Asset is also intended to measure changes in needs and risk of reoffending over time. The interim report on the validity and reliability of ASSET (Roberts et al 2001) contained a detailed account of practitioners’ views on the purpose, design and use of ASSET. Feedback was obtained from staff in 39 YOTS using confidential questionnaires. Two hundred and thirteen completed questionnaires were received back from 350 practitioners and 42 out of 60 from operational managers during autumn 2000. In addition, group discussions were held with staff from 15 YOTS to allow for more in-depth discussion. Evaluation included a small-scale programme of qualitative interviews with YOT practitioners; to investigate the views of practitioners, to shed light on the context in which they work, and to provide a background to the administrative data collected in the JCS. Face-to-face interviews, following a common semi-

structured interview schedule were carried out with three to four practitioners in 28 of 30 YOTS taking part in the JCS.

Practitioners defined the purpose of the ASSET form as:

- for making a comprehensive and holistic assessment;
- for identifying the needs of a young person;
- for identifying factors contributing to offending behaviour;
- for identifying risk and vulnerability;
- for identifying positive factors as well as problems.

Few respondents referred explicitly to identifying factors linked to offending behaviour – most referred in more general terms to identifying needs. Ten percent of questionnaire respondents referred to ASSET as a tool for collecting statistical information, typically framed in a negative context with some recognition of the potential value of gathering such information.

Operational managers described the potential value of ASSET for:

- encouraging consistent practice,
- strengthening links between assessment and intervention planning and,
- informing decisions about resource allocation.

Training

In an evaluation in 2001 one-fifth of practitioner respondents indicated that they had not received any training on ASSET at all. Others had received some training, either through national Youth Justice Board funded events or through local managers and colleagues. Most teams expressed a wish for further training and the demand and need for training remained high. The Youth Justice Board funded training on 'Assessment skills, report writing and supervision planning' and provided subsequent training material for teams to use locally. Frequently cited reasons for low reapplication rates included a lack of time and problems with IT systems not being fully operational. In some teams, all staff completed ASSET regardless of their professional background whilst, in others, ASSET was seen as exclusively for those from a social work or probation background. Some people used ASSET as an interview schedule whilst others completed it after they had interviewed a young person and gathered information from a range of relevant sources.

Characteristics of assessment reports

Approximately 60% of those who responded to the practitioners' questionnaire thought that ASSET was relevant to the process of preparing presentencing reports. Twenty percent held the view that the PSR took priority and ASSET was either unhelpful or made no difference to the process or quality of their report writing. Experienced practitioners were more likely to view ASSET as unnecessary and some perceived it as an insult to their professional skills. There were concerns that ASSET might encourage a return to the style of Social Enquiry Reports given its inclusion of some social, health and welfare issues. This was linked to uncertainty about how much of the information within ASSET ought to be included in a PSR.

Staff recognised that PSRs are targeted towards a specific audience and that not everything within ASSET necessarily needed to be included in a report. There were some differences between the professional groups within YOTS as to how this information should be selected. For example, some police officers were critical of social workers for leaving out negative information about a young person. Practitioners were keen to see closer integration between ASSET and PSRs through, for example, the use of IT systems to transfer information from ASSET more directly.

Discussion groups revealed that practitioners felt that 'thinking and behaviour', 'attitudes to offending' and 'motivation to change' were considered the most useful sections of the ASSET tool, possibly because of potential links to Pre-sentencing reports. Certain sections were perceived as not detailed enough (e.g. statutory education and physical health). In other sections, such as emotional and mental health, practitioners sometimes felt that they were not qualified to make reliable assessments. A key finding was that staff emphasised the importance of explaining the concepts used within ASSET to young people in a way that they could understand. Particular professional skills are required to elicit information from a young person and his/her family in a way that was culturally and contextually sensitive.

While most workers accepted the usefulness of the structured approach to assessment that ASSET offered, practical difficulties around local policies, team structures, workload, individual working styles, resources and training, all affected the way that it was used in practice. Social care practitioners disliked using the detailed and time-consuming ASSET forms where cases involved minor offences. They were also wary about offering scaled numerical ratings in certain areas which were seen as highly subjective. They noted the need to adapt certain questions to suit each young person. Adaptation was especially important when dealing with young people from ethnic minorities. The most positively received part of the tool was the 'What do YOU

think?’ self-assessment valued because it offered an effective way of engaging the young person in discussion. Roberts et al (2001) report that practitioners often viewed the completion of the ASSET forms as isolated from other tasks such as report writing, reviewing progress or intervention planning (Roberts et al, 2001). Baker et al (2005) found that completed intervention plans often did not reflect the outcomes of the assessment. For example, ‘issues identified as being associated with a high risk of re-offending were not always incorporated into intervention plan targets... [and]... there appeared to be a tendency to create ‘standardised’ plans...’ (Baker et al, 2005: 4)

Electronic versions of the ASSET tool were thought to encourage greater sharing of ASSET information between team members-2 (Baker et al, 2005). However some team members-2 were reluctant to utilise information contained in ASSETs previously completed by other practitioners. This reluctance seemed to indicate that they were unwilling to trust the judgement of colleagues, preferring to start again with their own assessment. This observation can be linked to the need to create ownership of assessment documentation and the assessment process by all professionals.

Concerns about data protection and a feeling by YOT staff that other organisations did not know about or understand the purpose of ASSET were reported as barriers to inter-agency sharing. YOTS stated that they did send copies of ASSET to the secure estate for young people receiving custodial sentences, but expressed concern that these would not be read or that, if they were read, the information might be used inappropriately. Examples were also cited of other agencies being unwilling to accept referrals on the basis of ASSET.

Staff felt that ASSET was a useful tool by prompting discussion and thus providing an explicit opportunity for a young person to express their views, ASSET can also highlight areas of concern that an assessment might otherwise have missed. One practitioner summed up the positive feeling towards the form by describing it as “their voice”. A minority of staff felt that the form was not useful in that it did not provide any information over and above that which would be obtained through the normal interviewing process. Others acknowledged that it could be helpful but felt that they did not have enough time to use it when preparing a PSR or that it might be inappropriate to use in the early stages of contact with an offender given the personal nature of some of the questions. There was some criticism of the design of the form in that it was felt to be too ‘boring’ and also too complicated for some young people to understand.

The idea of IT based interactive versions of the form was broadly welcomed. Although most teams had an electronic version of ASSET in place at the time of the interim report (Roberts et al 2001), many had difficulties in integrating them into practice. Consequently, staff felt that their experiences of using electronic ASSET had not matched their expectations. A number of practitioners felt that they lacked basic IT skills and this made it difficult to use an electronic version of ASSET. Staff also found that the process of inputting ASSET onto IT systems could be very time consuming. The key benefits of electronic systems were that it became easier to share information within teams and to reapply ASSET for reviews or at the end of interventions. Difficulties included confusion caused by differences between the paper format and the on-screen layout of the electronic versions and the fact that some IT systems produce a print-out which bears little resemblance to the original ASSET design.

Participants were asked to outline the strengths and weaknesses of ASSET for identifying needs and risks. This was asked as an open question, with spontaneous answers subsequently coded into categories.

Strengths of ASSET

Strengths were considered to be, in particular, around providing a useful checklist (a framework), and that ASSET helps to focus thinking (and ensures that all relevant aspects of a young person's life are covered at an initial stage of contact with the young person). Aligned to this were the views that ASSET was comprehensive, and encouraged a holistic view of the young person's life.

Weaknesses of ASSET

Perceived weaknesses of ASSET mainly related to the process of completing the Asset assessment, rather than the issue of identifying needs and risks). These included: that it could be subjective; that it was poorly structured/repetitive and restrictive in its approach; also that it could be time-consuming/lengthy. Also, others remarked that not all of the questions were relevant to all young people, and that, despite its length, important areas that were not covered. Practitioners reported that it was difficult to obtain accurate information from the young person, that it led to so called "scoring confusion", and it was described as deficit-led and "not 'young person-friendly'".

ASSET was not considered particularly strong in identifying risk or need. It proved difficult to explore such issues as 'emotional and mental health', 'family and personal relationships' and 'perception of self and others' with young people. Difficulties with exploring 'emotional and

mental health' were sometimes affected by the limited skills of workers in this area. Issues around 'family and personal relationships' were reported to be prone to concealment by the young person, particularly with regard to domestic violence, abuse, family substance misuse and family offending. The presence of a parent in the interview was also seen as a compounding factor by some interviewees.

The ASSET areas interviewees found most difficult to address through their work related to 'family and personal relationships' and 'living arrangements'. One suggested reason why these were considered difficult relates to the fact that resolving these difficulties was not 'in the gift of the YOT' suggesting that co-terminosity of the assessment scope and the scope of professional practice might be considered helpful. Similarly other areas that proved problematic were those arising from structural difficulties, gaps in services, poor or overstretched services and poor partnerships.

Component tools (i.e. tools not specific to HSB that might form part of an overall assessment)

No references were identified that explored stakeholder views' of component tools that might form part of an overall assessment (e.g. Beck Depression Inventory, PCL-YV, Trauma Symptom Checklist, Multiphasic Sex Inventory, Beckett's ASAP measures, etc.)

DISCUSSION

Statement of principal findings

Question 1: What factors contribute to a successful assessment process for adolescents with harmful sexual behaviour?

This review has focused on the assessment process in the specific context of adolescents who may be at risk for harmful sexual behaviour. The evidence reveals that the assessment process should be “joined up” with subsequent therapeutic interventions. This increases client confidence in the value of the assessment process. It is also helpful to try to ensure some continuity between assessment and intervention. Assessment should be viewed as the first part of the treatment pathway with the potential to influence the subsequent engagement with a client. Particularly critical is seen the need for prompt assessment with delays or inadequate assessments leading to a risk to safety and a subsequent delay in accessing appropriate interventions or services.

Inter-agency divisions, particularly the separation and potential conflict between the requirements of the criminal justice and social work sectors are seen as unhelpful. This is particularly the case where delays in the judicial process have a knock on effect on assessment and hence delay referral, treatment or entry into some residential provision.

Lack of training, particularly among social work practitioners, is seen as a further potentially delaying factor. Assessment is seen by some to be the responsibility only of the Youth Offending Teams (YOTs), Reluctance was also expressed at sharing of assessments and at using the assessments performed by another. On the other hand social work practitioners might not question the accuracy or appropriateness of information contained within an existing assessment and therefore might lack professional ownership of that assessment.

It should be noted that social work research is populated by large numbers of research papers and book chapters that relate more generally to the assessment process. Such sources may offer a useful evidence base in implementing effective assessment processes (Platt, 2001; Spratt & Callan, 2004; Platt, 2006): Many of these wider sources highlight the dual focus of the assessment process on both helping and safeguarding (protecting).

Question 2: What factors inform our understanding of the value of specific tools or methods when assessing adolescents for harmful sexual behaviour?

The evidence base is very uneven in relation to the use of specific assessment tools. Two detailed evaluations of assessment approaches (AIM/AIM2 and ASSET) are UK based and contain rich data which might be thought to inform our understanding of these tools. However these evaluations are not formally constructed as research studies. As a consequence they may be subject to selection bias. Importantly they do not include any comparators. There is some evidence to suggest that an assessment tool offers a reproducible process and structure for initiating a dialogue with the adolescent and other informants. Focusing on the tool itself can downplay the importance of the therapeutic relationship and the safe environment which have already been found to be important in delivery of the therapeutic approaches. It can also minimise the requirements for managerial support and training.

Notwithstanding such cautions we can observe that the initiatives to develop the AIM2 assessment framework for multi-agency working and the ASSET tool within the criminal justice system has resulted in a perceived improvement in the availability of a more standardised approach. These particular tools also seem to have contributed to a heightened awareness of the importance of assessment more generally although awareness is by no means universal with many misperceptions persisting (Deacon, 2015).

Methodological considerations

The qualitative evidence synthesis has revealed an evidence base which is uneven, scarce and, generally, of only moderate quality. Partly this is attributable to difficulties in information retrieval – assessment is not typically identified within the titles and abstracts of potentially relevant qualitative reports and useful data can therefore only be identified from detailed examination of full-text and persistent follow up of cross-references. However it is noticeable that few authors have explored the assessment process in general and the existing evaluations of specific instruments or tools are opportunistic, purpose-specific and may be subject to selection bias.

Further research

In her recent thesis Deacon (2015) concludes that “direction is needed from policy makers as to what specific research and tools frontline social work practitioners should use”. Social care practitioners particularly need guidance on the identification of problematic sexualised

behaviour in children in order to develop their competence and confidence in this challenging area of work. She states that “it is imperative that the government commissions more research to provide these guidelines”. Citing the Brook Sexual Behaviours Traffic Light Tool (2012) she states that the area of tool development is under researched. Indeed we were unable to find any qualitative evaluations of this particular tool or indeed of many tools used for forensic or actuarial assessments as favoured in the United States (Miccio-Fonseca & Rasmussen, 2011). However Deacon (2015) also reminds us that assessment must not become another box ticking exercise, but should give practitioners a tool to use in conjunction with their professional judgement.

Conclusion

Practitioners need to be convinced of the value and relevance of assessment for their everyday practice if they are to complete assessment tools and use them effectively. For practitioners this may mean receiving regular feedback about data being collected and how this is being used to influence policy and practice (Deacon, 2015). For adolescents and their carers this may require that the assessments are conducted promptly and completely and that the resultant assessments are acted upon in selecting appropriate interventions and access to services. For organisations there is a need to use the assessment process as a way of delivering “joined up care” – integrating the assessment and therapeutic processes and facilitating work across agencies. Developments in electronic sharing of assessments are one potential route for achieving this instrumentally but there is substantive evidence to suggest that much remains to be achieved in terms of improved and shared understanding of the assessment process, training and managerial support.

Important components of an assessment of a child or adolescent with concerning or harmful sexual behaviour (this is not an exhaustive list but summarises components highlighted in the evidence review)

- Comprehensive history of the person
- A cultural perspective
- Standardized, appropriate and validated assessment measure that can be shared with other agencies
- Sensitivity to the sexual content of the questions which may be too explicit for some communities
- Opportunity to talk to other professionals and gather different types of information
- Multiagency
- Assessment of the harmful sexual behavior
- Post-assessment case conference formalized as part of the assessment
- Holistic – recognizing the significance of the family, child development, and peers
- Uses a tool/s that differentiates between experimentation and harmful sexual behavior that might lead to re-offence
- Should not be a static one off process
- Family support during the assessment
- Prioritization of high risk adolescents
- Young person kept involved and engaged

Features of the process of assessment that were regarded as important to effective assessment described in the evidence review

Relationship between young person, family and the professional/s

- Relationship between the young person and the professional are key to successful assessment

- Provide more information to parents and families before it is carried out
- Many express a preference for the assessment to take place in the young person's own home
- Friendly interactions with the professionals greatly ease anxiety
- Continuity between the practitioner involved in the assessment and the young person
- Rapport, trust, communication
- Recognize the young person may be both a 'victim' and an 'offender'
- Assessment is part of a therapeutic pathway
- Delayed or incomplete assessments are a barrier to effective interventions
- Assessment process – contributes a good working relationship between adolescents and social workers and parents

Continuity

- Earlier and more thorough assessment is beneficial
- Family and young person are able to see how the assessment is being used.

Multiagency

- Poor interagency working adversely affects assessment, particularly between social justice and social care sectors
- Tensions may arise between criminal justice and social care systems and procedures delaying assessments
- Tensions between criminal and therapeutic considerations

Assessment tool

- Professional, prompt, offered recommendations and well researched
- Electronic systems needed
- Training needed in use of tool
- Managerial support needed during process
- Should explicitly link assessment to service providers
- AIM2 and ASSET felt to positively assist in facilitating a more standardized approach

Harmful effects

- Recognize the process of initiating assessment may result in stigmatization

Barriers to effective assessment

- Can be a bottleneck to appropriate care
- Waiting list seem to exacerbate the problems regarding suboptimal delayed assessment
- Lack of training and lack of clarity over roles is perceived to be a barrier to sharing of assessments

References

Baker, K. (2004). Is Asset really an asset? Assessment of young offenders in practice. What Works in Probation and Youth Justice: Developing Evidence-Based Practice. Cullompton: Willan publishing.

Baker, K., Jones, S., Roberts, C., Merrington, S., Centre for Criminological Research, & United Kingdom. (2005). *Evaluation of the Validity and Reliability of the Youth Justice Board's Assessment for Young Offenders*.

Belton, E., Barnard, M. and Cotmore, R. (2014) Turn the page: learning from a manualised approach to treating harmful sexual behaviour. London: NSPCC.

Borum, R., Bartel, P., & Forth, A. (2003). Manual for the Structured Assessment of Violence Risk in Youth, Version 1.1. Tampa: University of South Florida

Brook. (2012) Sexual Behaviours Traffic Light Tool. A Guide to Identifying Sexual Behaviours. Available from: www.brook.org.uk/index.php/traffic-light-tool-0-to-5 [Accessed April 2014].

Campbell F; Booth A; Stepanova E; Hackett S; Sutton A; Hynes K; (2016) in press. Harmful sexual behaviour in children: Evidence for identifying and helping children and young people who display harmful sexual behaviour. NICE. London.

Cleaver, H. and Walker, S. (2004) 'From policy to practice: the implementation of a new framework for social work assessments of children and families', *Child and Family Social Work*, 9(1), 81–90.

Corby, B., Millar, M., & Pope, A. (2002). Assessing children in need assessments-a parental perspective. *Practice*, 14(4), 5-15.

Corby, B., Millar, M. and Pope, A. (2002b) Having Your Say: Parents' Experiences of Assessment Under the Framework for the Assessment of Children in Need and their Families, *Universities of Liverpool & Central Lancashire*.

Deacon, L A (2015) *Children's Social Care Services' Response to Children who Display Sexually Harmful Behaviour*. Doctoral thesis, Durham University.

Department of Health, Department for Education and Employment, Home Office (2000) Framework for the Assessment of Children in Need and Their Families, London: HMSO

Epperson DL & Ralston CA. Juvenile Sexual Offender Recidivism Assessment Tool-II (JSORRAT-II). Paper presented at the Meeting of the Association for the Treatment of Sexual Abusers . 2006.

Elkovitch, N., Viljoen, J.L., Scalora, M.J., & Ullman, D. 2008. Assessing risk of reoffending in adolescents who have committed a sexual offence: The accuracy of clinical judgments after completion of risk assessment instruments. *Behavioral Sciences and the Law*, 26, (4) 2008

Finkelhor, D. Child sexual abuse: New theory and research. New York: Free Press, 1984.

Geary, J. E. (2007). *Adolescent sexual offender treatment programmes in New Zealand: A process evaluation* (Doctoral dissertation, ResearchSpace@ Auckland).

Geary, J., Lambie, I., & Seymour, F. (2011). Consumer perspectives of New Zealand community treatment programmes for sexually abusive youth. *Journal of sexual aggression*, 17(2), 181-195. (4)

Griffin, H., & Beech, A. (2004). Evaluation of the AIM Framework for the assessment of adolescents who display sexually harmful behaviour. Available at: www.youth-justice-board.gov.uk (accessed 18

August 2015).

Griffin, H. L., Beech, A., Print, B., Bradshaw, H., & Quayle, J. (2008). The development and initial testing of the AIM2 framework to assess risk and strengths in young people who sexually offend. *Journal of sexual aggression*, 14 (3), 211-225.

Griffin, H.L. & Vettor, S. 2012. Predicting sexual re-offending in a UK sample of adolescents with intellectual disabilities. *Journal of sexual aggression*, 18, (1) 64-80
Grimshaw, R., Downes, P. G., & Smith, A. (2008). Young people who sexually abuse. Youth Justice Board. Retrieved August 19th 2015.

Hackett, S. (2004) *What Works for Children and Young People with Harmful Sexual Behaviours*. Essex: Barnardo's

Hackett, S. (2011). *Children and young people with harmful sexual behaviours. Children Behaving Badly: Peer Violence Between Children and Young People*. Chichester: Wiley/Blackwell, 121-135.

Hall S (2006) Children with harmful sexual behaviours—what promotes good practice? a study of one social services department. *Child Abuse Review* 15: 273–84. doi:10.1002/car.926.

Hall, S. C. (2010). Exploring implications and benefits of holistic working with young people who have sexually harmed others. De Montfort University, Leicester. [Thesis]
<https://v00dor00001d.dmu.ac.uk/handle/2086/4977>

Harris, N. (2011). Does responsive regulation offer an alternative? Questioning the role of formalistic assessment in child protection investigations. *British Journal of Social Work*, 41(7), 1383-1403.

Harris, N. (2012). Assessment: When does it help and when does it hinder? Parents' experiences of the assessment process. *Child & Family Social Work*, 17(2), 180-191.

Harrison, L. (2009). The ambiguity of juvenile Sexual offenders. *Internet Journal of Criminology* retrieved 02/09/2015 from http://www.Internetjournalofcriminology.com/Harrison_Juvenile_Sexual_Offenders_July_09.df.

Hayes, D., & Spratt, T. (2009). Child welfare interventions: Patterns of social work practice. *British Journal of Social Work*, 39(8), 1575-1597.

Horwath, J. (2011). See the practitioner, see the child: The Framework for the Assessment of Children in Need and their Families ten years on. *British Journal of Social Work*, 41(6), 1070-1087.

Ladwa-Thomas U, Sanders R (1999) Juvenile sex abusers: perceptions of social work practitioners. *Child Abuse Review* 8:55–62.

Martinez R, Flores J, & Rosenfeld B 2007. Validity of the juvenile sex offender assessment protocol - II (J-SOAP-II) in a sample of urban minority youth.
Criminal Justice and Behaviour, 34, (10) 1284-1295

Miccio-Fonseca, L. C. (2006). *Multiplex Empirically Guided Inventory of Ecological Aggregates for Assessing Sexually Abusive Children and young people (Ages 19 and Under) - MEGA*. San Diego, CA.

Miccio-Fonseca, L. C. (2009). *MEGA*: A new paradigm in protocol assessing sexually abusive children

and young people. *Journal of Child and Adolescent Trauma*, 2, 124-141.

Miccio-Fonseca, L. C. (2010). MEGAJ: An ecological risk assessment tool of risk and protective factors for assessing sexually abusive children and young people. *Journal of Aggression, Maltreatment, and Trauma*, 19, 734-756.

Miccio-Fonseca, L. C., & Rasmussen, L. A. (2011). A concise review on validated risk assessment tools for sexually abusive youth. *Sexual Offender Treatment*, 6(2), 1-13.

Millar, M., & Corby, B. (2006). The Framework for the Assessment of Children in Need and their Families—A Basis for a 'Therapeutic' Encounter?. *British Journal of Social Work*, 36(6), 887-899.

Morrison, T. and Henniker, J. (2006) 'Building a comprehensive inter-agency assessment and intervention system for young people who sexually harm: the AIM project' in Erooga, M. and Masson, H. (eds) *Children and Young People Who Sexually Abuse Others*. London: Routledge: 31 – 50

Platt, D. (2001). Refocusing children's services: evaluation of an initial assessment process. *Child and Family Social Work*, 6(2), 139-148.

Platt, D. (2006). Investigation or initial assessment of child concerns? The impact of the refocusing initiative on social work practice. *British Journal of Social Work*, 36(2), 267-281.

Prentky, R., & Righthand, S. (2003). *Juvenile sex offender assessment protocol-II (J-SOAP-II) manual*. Washington, DC: US Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Prentky RA 2006, *Risk Management of Sexually Abusive Youth: A Follow Up Study*
7, U.S. Department of Justice, 214261

Prentky, R., Harris, B., Frizzell, K., & Righthand, S. (2000). An actuarial procedure for assessing risk with juvenile sex offenders. *Sexual Abuse: A Journal of Research and Treatment*, 12(2), 72-93.

Rajlic A & Gretton H M 2010. An examination of two sexual recidivism risk measures in adolescent offender. The Moderating Effect of Offender Type
5. *Criminal Justice and Behaviour*, 37, 1066-1085

Roberts C, Baker K, Merrington S and Jones S (2001) *The validity and reliability of ASSET: interim report to the Youth Justice Board*, University of Oxford.

Ryan, G. (2000). Childhood sexuality: a decade of study. Part I—research and curriculum development. *Child abuse & neglect*, 24(1), 33-48.

Shemmings, Y. and Shemmings, D. (2001) *Empowering Children and Family Members-2 to Participate in the Assessment Process* in Horwath, J. (ed.) *The Child's World*. London: Jessica Kingsley Publishers Ltd.

Spratt, T., & Callan, J. (2004). Parents' views on social work interventions in child welfare cases. *The British Journal of Social Work*, 199-224.

Viljoen JL, Scalora M, Cuadra L, Bader S, Chavez V, Ullman D, & Lawrence L (2007) *Assessing risk for violence in adolescents who have sexually offended. A comparison of the J-SOAP-II, J-SORRAT-II and SAVRY*. *Criminal Justice and Behaviour*

Viljoen JL, Elkovitch N, Scalora MJ, & Ullman D (2009) *Assessment of reoffence risk in adolescents*

who have committed sexual offences. Predictive Validity of the ERASOR, PCL:YV, YLS/CMI and Static-99 Criminal Justice and Behaviour

Wilson, E., & Hinks, S. (2011). Assessing the predictive validity of the Asset youth risk assessment tool using the Juvenile Cohort Study (JCS). Ministry of Justice Research Series 10, 11.

Worling, J. R. (2004). The Estimate of Adolescent Sexual Offence Recidivism (ERASOR): Preliminary psychometric data. *Sexual Abuse: A Journal of Research and Treatment*, 16(3), 235-254.

Worling, J. R., & Curwen, T. (2001) The ERASOR: Estimate of Risk of Adolescent Sexual Offence Recidivism (Version 2.0). Toronto, Ontario, Canada: Sexual Abuse Family Education and Treatment.

Worling, J.R., Bookalam, D., & Litteljohn, A. (2012) Prospective validity of the Estimate of Risk of Adolescent Sexual Offence Recidivism (ERASOR). *Sexual Abuse: Journal of Research & Treatment*, 24, (3) 203-223

Worling, J.R. & Langton, C.M. (2015) A prospective investigation of factors that predict desistance from recidivism for adolescents who have sexually offended. *Sexual Abuse: Journal of Research & Treatment*, 27, (1) 127-142

Appendix 1 - Excluded studies with reasons

Study	Reason for exclusion
Albertini, G., Polito, E., Sara, M., Di, G.G., & Onorati, P. 2006. Compulsive masturbation in infantile autism treated by mirtazapine. <i>Pediatric Neurology</i> , 34, (5) 417-418	Case report of treatment of autistic boy with hypersexual behaviour
Aldernen, M. (2001) Juvenile Sex Offender Treatment Program Provides Residential, Aftercare Service, <i>On Good Authority</i> (Illinois Criminal Justice Information Authority) 4 (9): 1-4.	Not Assessment
Allan, J. (2006). Whose job is poverty? The problems of therapeutic intervention with children who are sexually violent. <i>Child Abuse Review</i> , 15(1), 55-70.	Not Assessment
Almond, T. (2012). Asset: An assessment tool that safeguards or stigmatizes young offenders?. <i>Probation Journal</i> , 59(2), 138-150.	Not Qualitative
Almond, T. J. (2014). Working with children and young people with harmful sexual behaviours: exploring impact on practitioners and sources of support. <i>Journal of sexual aggression</i> , 20(3), 333-353.	Not Assessment
Arkowitz, S. & Vess, J. 2003. An evaluation of the Bumby RAPE and MOLEST scales as measures of cognitive distortions with civilly committed sexual offenders. <i>Sexual Abuse: Journal of Research & Treatment</i> , 15, (4) 237-249 Ref ID: 142	Measuring cognitive distortions in sexual offenders
Babchishin, K.M., Nunes, K.L., & Kessous, N. 2014. A multimodal examination of sexual interest in children: a comparison of sex offenders and nonsex offenders. <i>Sexual Abuse: Journal of Research & Treatment</i> , 26, (4) 343-374 Ref ID: 185	Study examining the discriminative and convergent validity of three different measures of sexual interest in children.
Baumgartner, J.V., Scalora, M.J., & Huss, M.T. 2002. Assessment of the Wilson Sex Fantasy Questionnaire among child molesters and nonsexual forensic offenders. <i>Sexual Abuse: Journal of Research & Treatment</i> , 14, (1) 19-30 Ref ID: 267	Study with adult offenders.
Becerra-Garcia, J.A., Garcia-Leon, A., Muela-Martinez, J.A., & Egan, V. 2013. A controlled study of the Big Five personality dimensions in sex offenders, non-sex offenders and non-offenders: Relationship with offending behaviour and childhood abuse. <i>Journal of Forensic Psychiatry and Psychology</i> , 24, (2) April Ref ID: 274	Not assessment
Boswell, G.. and Wedge, P. (2002) Sexually Abusive Adolescent Males: An Evaluation of a Residential Therapeutic Facility, Community and Criminal Justice Monograph 3. Leicester and Bedford: De Montfort University.	Not Assessment
Burton, D.L. 2008. An exploratory evaluation of the contribution of personality and childhood sexual victimization to the development of sexually abusive behavior.	Assessed the MACI and the

Sexual Abuse: Journal of Research & Treatment, 20, (1) 102-115	CTQ tools to establish the contribution of personality to the development of sexually abusive behaviour.
Canales, D.D., Olver, M.E., & Wong, S.C. 2009. Construct validity of the Violence Risk Scale--Sexual Offender Version for measuring sexual deviance. Sexual Abuse: Journal of Research & Treatment, 21, (4) 474-492 Ref ID: 607	Study with adult sex offenders
Coxe, R. & Holmes, W. 2009. A comparative study of two groups of sex offenders identified as high and low risk on the static-99. Journal of Child Sexual Abuse, 18, (2) 137-153 Ref ID: 798	
Craig, L.A., Browne, K.D., & Stringer, I. 2004. Comparing sex offender risk assessment measures on a UK sample. International Journal of Offender Therapy & Comparative Criminology, 48, (1) 7-27 Ref ID: 805	Age range 13-61. Mean age 36.3 years
Craissati, J. & Beech, A. 2005. Risk prediction and failure in a complete urban sample of sex offenders. Journal of Forensic Psychiatry and Psychology, 16, (1) March Ref ID: 810	Risk prediction in adult sex offenders
Criminal Justice Joint Inspection (2013) Examining Multi-agency Responses to Children and Young People who Sexually Offend: A Joint Inspection of the Effectiveness of Multi-agency Work with Children and Young People in England and Wales who have Committed Sexual Offences and were Supervised in the Community. England and Wales: Criminal Justice Joint Inspection.	Not Qualitative Research
Daniel, M. R., Weir, S., & Tiffin, P. A. (2013). A novel CAMHS risk assessment system: clinicians' views. <i>The Journal of Forensic Practice</i> , 15(3), 182-191.	Not sexual harm focus
Ducro, C. & Pham, T. 2006. Evaluation of the SORAG and the Static-99 on Belgian sex offenders committed to a forensic facility. Sexual Abuse: Journal of Research & Treatment, 18, (1) 15-26 Ref ID: 1000	Study evaluating the SORAG and the Static-99 in adult sex offenders.
Eliasov, N. (2004) Evaluating the SAYStop Diversion Programme: Findings from the third follow-up study, June-Nov 2003, South African Young Sex Offenders Project: a diversion programme for young sex offenders. Cape Town: Cape Town University	Not Assessment
Faller KC, Cordisco-Steele L, Nelson-Gardell D. Allegations of sexual abuse of a child: what to do when a single forensic interview isn't enough. Journal of Child Sexual Abuse 2010 Sep;19(5):572-89.	Not Assessment
Fernandez, Y.M., Marshall, W.L., Lightbody, S., & O'Sullivan, C. 1999. The child molester empathy measure: description and examination of its reliability and validity. Sexual Abuse: Journal of Research & Treatment, 11, (1) 17-31 Ref ID: 1169	A study to refine a measure of victim empathy – adult participants.
Fortune, C.A. & Lambie, I. 2006. Sexually abusive youth: A review of recidivism studies and methodological issues for future research. Clinical Psychology Review,	Review article

26, (8) December

- Franey, K.C., Viglione, D.J., Wayson, P., Clipson, C. and Brager, R. (2004) An Investigation of Successfully Treated Adolescent Sex Offenders, *Journal of Child Sexual Abuse*, 13, 295-317. Not Assessment
- Grady, M.D. & Rose, R.A. 2011. The Empathy Index: an evaluation of the psychometric properties of a new empathy measure for sex offenders. *Journal of Interpersonal Violence*, 26, (18) 3790-3814
Ref ID: 1467 Study designed to measure empathy in sex offenders.
- Green, L. and Masson, H. (2002) Adolescents Who Sexually Abuse and Residential Accommodation: Issues of Risk and Vulnerability, *British Journal of Social Work*, 32, 149-168. Not Assessment
- Hackett, S., Balfé, M., Masson, H., & Phillips, J. (2014). Family responses to young people who have sexually abused: anger, ambivalence and acceptance. *Children & Society*, 28(2), 128-139. Not Assessment
- Hackett, S., Masson, H. and Phillips, S. (2003) Mapping and Exploring Services for Young People who have Sexually Abused Others, A two-year research project funded by Youth Justice Board, NSPCC and The National Organisation for the Treatment of Abusers. University of Durham, University of Huddersfield: Centre for Applied Social and Community Studies, also the Centre for Applied Childhood Studies. Not Assessment
- Hackett, S., Masson, H., & Phillips, S. (2005). Services for young people who sexually abuse. *London: Youth Justice Board*. Survey with no qualitative data
- Hackett, S., Masson, H., & Phillips, S. (2006). Exploring consensus in practice with youth who are sexually abusive: Findings from a Delphi study of practitioner views in the United Kingdom and the Republic of Ireland. *Child maltreatment*, 11(2), 146-156. Not Assessment
- Hall, G.C., Graham, J.R., & Shepherd, J.B. 1991. Three methods of developing MMPI taxonomies of sexual offenders. *Journal of Personality Assessment*, 56, (1) 2-13
Ref ID: 1568
- Hall, G.C., Shepherd, J.B., & Mudrak, P. 1992. MMPI taxonomies of child sexual and nonsexual offenders: a cross-validation and extension. *Journal of Personality Assessment*, 58, (1) 127-137
Ref ID: 1567
- Hennessy, M., Walter, J.S., & Vess, J. 2002. An evaluation of the Empat as a measure of victim empathy with civilly committed sexual offenders. *Sexual Abuse: Journal of Research & Treatment*, 14, (3) 241-251
Ref ID: 1692 Study with adults
- Henniker, J., & Morrison, T. (2002). An Inter-agency Assessment Framework for Young People Who Sexually Abuse: Principles, Processes and Practicalities. *Child Care in Practice*, 8(2), 114-126. No qualitative data
- Hutchinson, J., Lovell, A., & Mason, T. (2012). Managing risk: a qualitative study of community-based professionals working with learning-disabled sex offenders. *Journal of psychiatric and mental health nursing*, 19(1), 53-61. Not Adolescents
- Ilacqua, G.E., Coulson, G.E., Lombardo, D., & Nutbrown, V. 1999. Predictive validity of the Young Offender Level of Service Inventory for criminal recidivism of male and female young offenders. *Psychological Reports*, 84, (3:Pt 2) t-8 Looks at prediction of criminal

Ref ID: 1860	recidivism using the YO-LSI tool
Kimonis, E.R., Fanniff, A., Borum, R., & Elliott, K. 2011. Clinician's perceptions of indicators of amenability to sex offender-specific treatment in juveniles. <i>Sexual Abuse: Journal of Research & Treatment</i> , 23, (2) 193-211 Ref ID: 2095	Study exploring clinicians perceptions of amenability to treatment in juvenile sex offenders.
Kolton, D.J., Boer, A., & Boer, D.P. 2001. A revision of the Abel and Becker Cognition Scale for intellectually disabled sexual offenders. <i>Sexual Abuse: Journal of Research & Treatment</i> , 13, (3) 217-219 Ref ID: 2157	Assessment of the readability of the Abel and Becker Cognition Scale
Langstrom, N. & Grann, M. 2002. Psychopathy and violent recidivism among young criminal offenders. <i>Acta Psychiatrica Scandinavica, Supplementum</i> .(412):86-92, 2002. (412) 86-92 Ref ID: 2269	Not assesment
Lawson, L. (2003) 'Becoming a Success Story: How Boys Who Have Molested Children Talk About Treatment', <i>Journal of Psychiatric and Mental Health Nursing</i> , 10, 259-268.	Not Assessment
Longo, R.E. 2005. An integrated experiential approach to treating young people who sexually abuse. <i>Journal of Child Sexual Abuse</i> , 13, (3-4) 2005 Ref ID: 2441	Intervention study
Losada-Paisey, G. 1998. Use of the MMPI-A to assess personality of juvenile male delinquents who are sex offenders and nonsex offenders. <i>Psychological Reports</i> , 83, (1) 115-122 Ref ID: 2448	Study examining relationship between personality and offender type
Masson, H., & Hackett, S. (2003). A decade on from the NCH Report (1992): Adolescent sexual aggression policy, practice and service delivery across the UK and Republic of Ireland. <i>Journal of Sexual Aggression</i> , 9(2), 109-124.	Not Qualitative
Masson, H., & Hackett, S. (2004). The extent and nature of work with adolescents who have sexual behaviour problems: findings from a survey of Youth Offending Teams in England and Wales. <i>Youth Justice</i> , 4(3), 160-177.	No Qualitative data
Masson, H., Hackett, S., Phillips, J., & Balfe, M. (2014). Looking Back on the Long-Term Fostering and Adoption of Children with Harmful Sexual Behaviours: Carers' Reflections on Their Experiences. <i>British Journal of Social Work</i> , 44(8), 2200-2217.	Not Assessment
Meys, U. (2006). The development and implementation of an intervention programme for young sexual offenders. University of Cape Town. [Thesis]	Not Assessment
Myers, S. (2007). (De) constructing the risk categories in the aim assessment model for children with sexually harmful behaviour. <i>Children & society</i> , 21(5), 365-377.	No Qualitative data
Myers, S., Eggleston, C. and Smoak, P. (2003) A Media Violence-Inspired Juvenile Sexual Homicide Offender 13 Years Later, <i>Journal of Forensic Sciences</i> , 48, 1-6.	Not Assessment
Rich, P. (2003). Understanding, assessing, and rehabilitating juvenile sexual offenders. New Jersey: John Wiley & Sons	Item Unavailable

Richardson, G., Kelly, T.P., Graham, F., & Bhate, S.R. 2004. A personality-based taxonomy of sexually abusive adolescents derived from the Millon Adolescent Clinical Inventory (MACI). <i>British Journal of Clinical Psychology</i> , 43, (Pt:3) 3-98	Study to classify a sample of sexually abusive adolescents on the basis of their personality and psychopathology scores obtained from the MACI tool.
Scheela, R.A. (2001) Sex Offender Treatment: Therapists' Experiences and Perceptions, <i>Issues in Mental Health and Nursing</i> , 22, 749-767.	Not Assessment
Shepherd, B. (2012). 'Asset: An assessment tool that safeguards or stigmatizes young offenders?' A practitioner response. <i>Probation Journal</i> , 59(3), 269-274.	Not Research
Slattery, P., Cherry, J., Swift, A., Tallon, M., & Doyle, I. (2012). From custody to community: Development of assessment and treatment for juveniles serving sentences for sex offences in an Irish context. <i>Journal of Sexual Aggression</i> , 18(1), 81-90.	No relevant qualitative data
Smith, C., Bradbury-Jones, C., Lazenbatt, A., & Taylor, J. (2013). Provision for young people who have displayed harmful sexual behaviour.	No qualitative data
SORENSEN, E. & Johnson, E. 1996. Subtypes of incarcerated delinquents constructed via cluster analysis. <i>Journal of Child Psychology & Psychiatry & Allied Disciplines</i> , 37, (3) 293-303	A study identifying homogeneous subtypes of delinquent adolescents using cluster analysis.
Steen, S. (2001) Contested Portrayals: Medical and Legal Social Control of Juvenile Sex Offenders, <i>Sociology Quarterly</i> , 42, 325-350.	Not Assessment
Tiffin, P. A., Kitchen, C. E., & Weir, S. (2015). Innovations in Practice: Piloting a new child and adolescent risk assessment suite in the United Kingdom. <i>Child and Adolescent Mental Health</i> .	Not Sexual Harm focus
Tildesley, E.A., Hops, H., Ary, D., & Andrews, J.A. 1995. Multitrait-multimethod model of adolescent deviance, drug use, academic, and sexual behaviors. <i>Journal of Psychopathology and Behavioral Assessment</i> , 17, (2) 1995	Study testing problem behaviour theory
Urbaniok, F., Endrass, J., Rossegger, A., & Noll, T. 2007. Violent and sexual offences: a validation of the predictive quality of the PCL:SV in Switzerland. <i>International Journal of Law & Psychiatry</i> , 30, (2) 147-152	Adult offenders
Urbaniok, F., Noll, T., Grunewald, S., Steinbach, J., & Endrass, J. 2006. Prediction of violent and sexual offences: A replication study of the VRAG in Switzerland. <i>Journal of Forensic Psychiatry and Psychology</i> , 17, (1) March	
van, V.E., Asscher, J., Stams, G.J., Hendriks, J., Bijleveld, C., & van der Laan, P. 2011. Moral judgment of young sex offenders with and without intellectual disabilities. <i>Research in Developmental Disabilities</i> , 32, (6) 2841-2846	Exclude – looking at questions to elicit information regarding moral judgement.
van, V.E., Hendriks, J., Stams, G.J., van, E.F., Bijleveld, C., van der Laan, P., & Asscher, J. 2011. Moral judgment, cognitive distortions and implicit theories in young sex	Exclude – looking at questions to

offenders. <i>Journal of Forensic Psychiatry and Psychology</i> , 22, (4) August	elicit information regarding moral judgement.
Vertommen, T., Toftegaard-Stoeckel, J., Vandevivere, L., Van Den Eede, F., & De Martelaer, K. (2014). A green flag for the Flag System? Towards a child protection policy in Flemish sport. <i>International Journal of Sport Policy and Politics</i> , (ahead-of-print), 1-17.	No qualitative data
Vick J, McRoy R, Matthews BM. Young female sex offenders: assessment and treatment issues. <i>Journal of Child Sexual Abuse</i> 2002;11(2):1-24.	Item not available
Vitacco, M.J., Caldwell, M., Ryba, N.L., Malesky, A., & Kurus, S.J. 2009. Assessing risk in adolescent sexual offenders: Recommendations for clinical practice. <i>Behavioral Sciences and the Law</i> , 27, (6) November-December Ref ID: 4193	Clinical review paper
Vosmer, S., Hackett, S., & Callanan, M. (2009). "Normal" and "inappropriate" childhood sexual behaviours: Findings from a Delphi study of professionals in the United Kingdom. <i>Journal of sexual aggression</i> , 15(3), 275-288.	Not Assessment
Watson, T. & Vess, J. 2007. Risk assessment of child-victim sex offenders for extended supervision in New Zealand. <i>Journal of Forensic Psychiatry and Psychology</i> , 18, (2) June Ref ID: 4268	Risk assessment of adult male child-victim sex offenders.
Wherry, J.N., Berres, A.K., Sim, L., & Friedrich, W.N. 2009. Factor structure of the Adolescent Clinical Sexual Behavior Inventory. <i>Journal of Child Sexual Abuse</i> , 18, (3) 233-246 Ref ID: 4296	This did not test predictive validity of the Adolexcent Clinical Sexual Behaviour Inventory-Self Report - but was designed to determine if it conformed to the five factor scale format that was previously used.
White, S.F., Cruise, K.R., & Frick, P.J. 2009. Differential correlates to self-report and parent-report of callous-unemotional traits in a sample of juvenile sexual offenders. <i>Behavioral Sciences and the Law</i> , 27, (6) November-December Ref ID: 4308	
Witt, P.H., Delrusso, J., Oppenheim, J., & Ferguson, G. 1996. Sex offender risk assessment and the law. <i>Journal of Psychiatry and Law</i> , 24, (3) Fall Ref ID: 4360	General review article
Wunsch, K. & Hasler, F. 2011. Gender differences in problematic childhood sexual behavior. European Child and Adolescent Psychiatry.Conference: 14th International Congress of ESCAP European Society for Child and Adolescent Psychiatry Helsinki Finland.Conference Start: 20110611 Conference End: 20110615.Conference Publication: (var.pagings), 20, (pp S176) June	
Wunsch, K. & Hasler, F. 2011. Sexual behavior and psychiatric symptoms in children. European Child and Adolescent Psychiatry.Conference: 14th International Congress of ESCAP European Society for Child and Adolescent Psychiatry Helsinki	Not available

Finland.Conference Start: 20110611 Conference End: 20110615.Conference
Publication: (var.pagings), 20, (pp S20) June

- | | |
|--|---|
| Wylie, L. A., & Griffin, H. L. (2013). G-map's application of the Good Lives Model to adolescent males who sexually harm: A case study. <i>Journal of sexual aggression, 19</i> (3), 345-356. | No Qualitative data |
| Yamamoto, B. A., & Kitano, N. (2015). Designing a safeguarding tool for Japanese professionals to identify, understand and respond to adolescent sexual behaviours. <i>Environmental health and preventive medicine, 20</i> (2), 141-145. | No Qualitative data |
| Yates, P., Allardyce, S., & MacQueen, S. (2012). Children who display harmful sexual behaviour: Assessing the risks of boys abusing at home, in the community or across both settings. <i>Journal of sexual aggression, 18</i> (1), 23-35. | Not Qualitative |
| Yeater, E.A., Lenberg, K.L., & Bryan, A.D. 2012. Predictors of sexual aggression among male juvenile offenders. <i>Journal of Interpersonal Violence, 27</i> , (7) 1242-1258 | Exploration of risk factors for sexual aggression |
| Yoder, J.R. 2014. Service approaches for youths who commit sexual crimes: a call for family-oriented models. <i>Journal of Evidence-Based Social Work, 11</i> , (4) 360-372 | Review article. |

Appendix 2 - Sample Search Strategy (from Ovid MEDLINE)

Population Terms

- 1 (sex* adj2 (harm* or risk* or abus* or agress* or unacceptable or offen* or force* or impos* or overly or coer* or inappropriate* or manipulat* or stigma* or shame or victim* or danger* or threat* or assault* or pressure* or violent or violence)).ti,ab.
- 2 (problem* adj2 sex* adj2 (behavio?r* or conduct*)).ti,ab.
- 3 *Sex Offences/
- 4 *Rape/
- 5 (rape or rapist).ti,ab.
- 6 *Unsafe Sex/
- 7 (unsafe adj2 sex).ti,ab.
- 8 or/1-7
- 9 (harm* or unacceptable or force* or impos* or coer* or inappropriate* or danger* or threat* or assault* or pressure* or violent or violence).ti,ab.
- 10 *Sexual Behavior/
- 11 (coitus or sexual intercourse).ti,ab.
- 12 (penetrat* adj2 sex).ti,ab.
- 13 *Coitus/
- 14 (masturbat* or self stimulat\$).ti,ab.
- 15 *Masturbation/
- 16 (sexual interaction or sexual exploration).ti,ab.
- 17 or/10-16
- 18 9 and 17

- 19 inappropriate touching.ti,ab.
- 20 (harm* or unacceptable or innappropraite*).ti,ab.
- 21 ((sexual* adj3 (swear* or word* or phrase* or slang or jargon)) or sexual* explicit).ti,ab.
- 22 20 and 21
- 23 sexting.ti,ab.
- 24 ((sex* or nud*) adj2 (message* or image* or picture* or photo*)).ti,ab.
- 25 23 or 24
- 26 8 or 18 or 19 or 22 or 25
- 27 *Child/
- 28 (child* or girl* or boy*).ti,ab.
- 29 (young people or young person* or young wom?n or young m?n or young female* or young male* or young adult* or youth*).ti,ab.
- 30 *Young Adult/
- 31 *Adolescent/
- 32 (adolescen* or teenage*).ti,ab.
- 33 Juvenile Delinquency/
- 34 delinquen*.ti,ab.
- 35 *Minors/
- 36 (minor or minors).ti,ab.
- 37 *Schools/
- 38 school*.ti,ab.
- 39 *"Latency Period (Psychology)"/
- 40 *Child, Preschool/

41 (preschool* or pre-school*).ti,ab.

42 (infant* or toddler* or youngster* or early adult* or kid or kids or underage or under age or teen* or offspring* or juvenile* or student*).ti,ab.

43 or/27-42

44 26 and 43

The above population terms were combined with a short filter for qualitative studies.

Appendix 3 – Summary of assessment tools

AIM2

The AIM framework (Print et al 2001) was the first to develop an integrated risks and strengths assessment, where individuals are rated with either high or low strengths and concerns. This tool was intended to provide a tool to assist practitioners to identify risks and needs and assist them with their initial decision-making. The original AIM assessment framework was introduced as a tool for use within Greater Manchester but it has since been adopted by a significant number of local authorities within the United Kingdom (Hackett et al 2005). In 2007 the original framework was updated, giving rise to AIM2, which incorporated recommendations from a small-scale evaluation (Griffin & Beech 2004) and empirical findings and clinical insights gained from the rollout of AIM. AIM2 was designed to assist with early stage assessments of young men of mainstream educational ability, aged between 12 and 18 years who are known to have sexually abused others.

AIM2 is based on an approach that assesses the static, stable dynamic, acute dynamic and trigger factors that lead to young people sexually abusing others. It contains 75 items and all items are used collectively to create one holistic assessment of the young person's risk and needs. The assessment uses a scoring system that assigns a differential weighing to items. The weighting reflects the extent to which the item is underpinned by credible research.

BERS-2 and BERS-2-2

No differences between the two versions of the BERS-2 with respect to the test items or scales that were examined in the Worling (2015) study. The second edition of the BERS-2 has a 5 item scale that is focused on the youth's career plans, and this was the only change to the item content of the BERS-2. There are five scales that are derived from the 52 items on the BERS-2. The Interpersonal Strength scale is based on 15 items and it is designed to measure the youth's ability to control their emotions and behaviours and to be respectful of others. Items on this scale are focused on issues such as respect for others, accepting responsibility and the appropriate expression of anger. The Family Involvement scale measures the youth's level of involvement with family members and the overall quality of family relationships. The third BERS-2 scale is Intrapersonal Strength. The 11 items that form this scale are designed to measure the youth's view of their personal competence and outlook on life. The final scale is Affective Strength. The 7 items on this scale measure capacity for emotional intimacy. With a

focus on the youth's ability to give and receive affection, express feelings, show concern for others, accept closeness and acknowledge painful feelings.

J-SOAP-II

Designed for use with adolescents aged 12 to 18.

“users should have training and experience in assessing juveniles who commit sexual offences and risk assessment in general, particularly as it pertains to juvenile sex offending It's an open assessment test, suitable for all relevant professionals irrespective of their training (i.e. social workers, probation officers, psychologists, etc.

Dynamic variables.

J-SOAP-II is a 28 item checklist whose purpose is to aid in the SR of risk factors that have been identified as being associated with sexual and criminal offending. Intended for use for males between the ages of 12 and 18 years who have a history of sexual offending or sexually coercive behaviour. It is important to note that the J-SOAP-II does not distinguish between risk of sexual re-offence and risk of general, criminal re-offence. The instrument contains four subscales: Sexual Drive/Preoccupation, Impulsive/Antisocial Behaviour, Intervention and Community Stability/Adjustment. Items on the J-SOAP-II are rated on a three point scale (absent, possibly present, clearly present) with a higher score representing greater risk. While total scores are obtained by summing the items on the four scales, at the present time the authors state that cut-off scores should not be used and the J-SOAP-II should function as an empirically informed guide rather than an actuarial tool. .

Estimate of Risk of Adolescent Sexual Offence Recidivism (ERASOR)

used to determine if protective factors for adolescent sexual recidivism could be identified in a prospective investigation. The ERASOR is a 25-item, single-scale instrument designed to structure professional judgments regarding the risk of sexual reoffending by youth aged 12 to 18. It includes extensive interviewing with the youth, the completion of a number of tests and questionnaires, document review, and meetings with caregivers. The tool is used among youth offenders convicted of and/or acknowledged criminal sexual behaviour(s) placed in residential care facilities. Among limitations identified is the small sample size due to relatively small numbers-2 of sexual and nonsexual recidivists. Another limitation was the absence of precise dates of sexual recidivism which would enable the clinicians to determine at-risk periods.

SAVRY

Designed for use with adolescents aged 12 to 18.

Need training in clinical forensic psychology to complete

Dynamic variables

The Structured Assessment of Violence Risk in Youth (SAVRY) is a guide for assessing risk for general violence in adolescents. Although it does not specifically assess risk for sexual offending, it includes a sexual violence in its definition of violence. The SAVRY is based on the structured professional judgment model of risk assessment and is composed of 24 risk items. As well as six protective factors. These items make up three sets of risk factors: Historical Risk Factors, Social and Contextual Risk Factors, and Individual Risk Factors. Each risk factor is coded as 'high', meaning a youth is high risk on this item, 'moderate' or 'low'. The SAVRY also assesses whether six protective factors are 'present' or 'absent'. Consistent with the structured professional judgment model, clinicians are to use item scores on risk and protective factors in order to make a professional judgment about a youth's risk for violence.

SAVRY is an instrument designed to assess risk of violence applied on residential treatment program among youth with intellectual and adaptive functioning at least at the borderline level, adjudicated delinquent of a sexual offence and mandated to receive treatment, and demonstrated self-control that would allow functioning in an open, unlocked treatment program. The major limitation is the difficulty to identify case-specific factors predicting sexual recidivism. Furthermore, small numbers of sexual reoffending hinder the opportunity to make accurate prediction of reoffending. The studies using SAVRY (Elkovitch et al., 2008) relied solely on file information to make violence risk decisions where risk and confidence judgments were made by Masters-level graduate student raters. Finally the examination of factors predictive of instrument-informed judgments of risk was limited to J-SOAP-II and SAVRY subscales.

J-SORRATT-II

Designed for use with adolescents aged 12 to 18.

Does NOT require clinical training to complete.

No dynamic variables

The J-SORRAT-II is a 12 item actuarial tool designed for assessing risk of violence among male juvenile offenders who were 12 to 18 years old at the time of their index sexual offence. A

number of items on the J-SORRAT-II focus on the youths sexual and nonsexual offence history (e.g. number of adjudications as a sex offender, number of victims in sex offences). Other variables examine youths treatment history (i.e. completion of sex offender treatment), school history and past victimization experiences. The J-SORRAT-II has an objective, criterion based scoring system. A number of J-SORRAT-II items are scored '0' or '1' to indicate whether the risk factor is present or absent. Other items are scored on a 3-point scale (0 to 2) or a 4 point scale (0 to 3) to indicate varying degrees of severity for a particular risk factor. The J-SORRAT-II was developed by identifying key predictors of sexual offending in a sample of 636 male youths who were adjudicated for a sex offence.

YLS/CMI

Designed to assess broad types of reoffending.

PCL:YV

The Hare Psychopathy Checklist: Youth Version is also frequently used in adolescent risk assessments. However, significant concerns have been raised regarding the assessment of psychopathy in youth.

Static-99

An adult tool that may be useful in examining adolescent samples. The Static-99 is a widely used and well-validated actuarial tool. The coding guidelines state that although the instrument is not intended to be used with individuals younger than 18, there may be some situations in which it could be used with adolescents, such as when offences occurred within the offender was 16 or 17 and the offences appear 'adult' in nature (Harris et al 2003).

Appendix 4 Evidence Tables – Quantitative Review

Study Setting Objective Design	Population Recruitment N Age Gender Ethnicity HSB	Study Design Selection Comparability Outcome	Test	Outcomes	Comments
<p>Griffin & Vettor (2012) UK</p> <p>To examine the ability of AIM2 and the adapted AIM assessment to predict sexual re-offending by adolescents with intellectual disabilities.</p>	<p>46 included Mean age of 15.3 years (SD = 1.4), range = 12-18 years)</p> <p>All males</p> <p>Recruitment Sampling frame of 317 males, aged 12-18 years with intellectual disability (ID)</p> <p>Selected from cases referred to G- MAP, an independent provider specialising in the assessment and treatment of adolescents who have displayed sexually harmful behaviour.</p> <p>Ethnicity White British 35/46 (76%) Mixed white and black Caribbean 1/46 (2%)</p>	<p>Information from the adolescent's case file was used to complete the AIM2 and adapted AIM assessment models. Only information provided at the initial assessment was used – increasing the validity of the study through replicating the quantity and quality of information available to case workers at the initial assessment stage.</p>	<p>AIM2 – consists of 75 assessment items: 26 static concerns factors, six static strengths factors, 25 dynamic concerns factors and 18 dynamic strengths factors.</p> <p>The adapted AIM for adolescents with ID – this contains 101 assessment items, divided into concerns and strengths factors. The items exist on a continuum separated in to, high, medium and low.</p>	<p>Re-offending data data: Mean follow-up period for re-offence data was 6 years (SD = 3.1, range 2-15.6 years). Re-offending was defined as a further commission of a sexually harmful incident, regardless of whether the initial sexual incident of the subsequent incident resulted in a charge or conviction. Criminal histories information was used to identify sexual re-offending and subsequent non-sexual offending was obtained from the police national computer and therefore these data had resulted from a charge or conviction.</p>	<p>Due to the retrospective nature of the study, many dynamic factors were difficult to gauge and would, in reality, have been more accessible to case workers at the time of undertaking the initial assessment.</p> <p>A limitation of this method is that files have been compiled for purposes relevant to the workers practice and organisational needs rather than for scientific research. Consequently file information ncan present a biased view, be incomplete and be inconsistent, this making it less effective as a research tool.</p>

	<p>Mixed white and Asian 1/46 (2%) Asian Pakistani 1/46 (2%) Asian other 1/46 (2%) Not known 7/46 (7%)</p> <p>Degree of ID Severe 2/46 (4%) Moderate 10/46 (22%) Mild 24/46 (52%) Not defined 10/46 (22%)</p> <p>Type of referral behaviour Contact sexual behaviours 40/46 (87%) Non-contact sexual behaviours 2/46 (4%) Both contact and non- contact sexual behaviours 4/46 (9%)</p> <p>.</p>			<p>Also some cases, informed by up-to- date information from relevant professionals.</p> <p>Sexual re-offending 9/46 (20%)</p> <p>Non-sexual reoffending 18/46 (39%) No known future offending 19/46 (41%)</p> <p>Adapted AIM - total concerns score AUC 0.78 p value 0.01 SE 0.08 95%CI 0.62 to 0.94</p> <p>Adapted AIM total strengths score AUC 0.51, p value 0.97, SE 0.12, 95% CI 0.28 to 0.73</p> <p>AIM2 total concerns score AUC 0.78, p-value 0.01, SE 0.08, 95% CI 0.63 to 0.94</p> <p>AIM2 total strengths score</p>	
--	--	--	--	---	--

				<p>AUC 0.29, p value 0.05, SE 0.10, 95% CI 0.09 to 0.49</p> <p>AIM2 combined concerns and strengths score AUC 0.79, p-value 0.01, SE 0.09, 95% CI 0.62 to 0.96</p>	
<p>Griffin et al (2008) UK</p> <p>To refine the AIM2 by:1) identifying and removing items that were either non-discriminatory or highly correlated; 2) gain practitioner feedback on the assessment tool; 3) identify items that were more ambiguous so that they could be redefined; and 4) use a small scale pilot study to help assess items that were more predictive of recidivism.</p>	<p>Recruitment *a subset of the above study larger cohort</p> <p>Criminal histories for 317 young people who had been referred historically to G-MAP services, were searched using the policy national computer – providing a national database for England and Wales.</p> <p>n 70 young people selected randomly.</p> <p>Mean age: 15.5 years (range 12-18 years)</p> <p>Ethnicity: White British 65/70 (80%) African Caribbean 1/70</p>	<p>Retrospective design. Once recidivism data collected, young person's cases were excluded where their identity could not be matched confidently and where details of criminal histories could not be interpreted fully due to the use of different coding.</p>	<p>For these 70 cases the AIM2 assessment was completed using the young person's file information. Only information applicable to an initial assessment was used, for example pre-sentence reports, historical psychology reports and referral information.</p>	<p>Recidivism was defined as a further incident of a sexual nature, which involved physical contact with the victim and resulted in a caution or conviction.</p> <p>Period of follow-up was recidivist data was mean 6 years (minimum 2 years).</p> <p>Risk of recidivism Concerns scale AUC 0.98, p<0.00001, CI 0.98 to 1.01</p> <p>Strengths scale AUC 0.94 p=<0.00001 CI 0.89 to 1</p>	<p>Same study as Griffin & Vettor (2012) – different subsets.</p>

	<p>(1%) Mixed race 1/70 (1%) Other 1/70 (1%) Unknown 11/70 (11%)</p> <p>HSB Indecent assault 42/70 *(60%) Rape 11/70 (16%) Escalation of sexualized behaviours 6/70 (3%) Gross indecency 2/700 (3%) Attempted rape 1/70 (1%) Unlawful sexual intercourse 1/70 (1%)</p> <p>These incidents involved sexual abuse against a combination of males, females, children, peers and adults.</p> <p>Excluded: where file information was sparse, or the young person had a low IQ,</p>				
<p>Worling et al (2012) Canada</p> <p>To examine the prospective validity of the ERASOR.</p>	<p>191 adolescent males Assessed between Jan 2001-Oct 2007 at 5 agencies in Ontario, Canada.</p> <p>12-19 years (M= 15.34;</p>	<p>Cohort study</p>	<p>Assessed by one of 22 clinicians (including the first author). Each assessment involved a combination of clinical interviews, psychological testing, document reviews</p>	<p>Follow up period: Ranged from 0.1 to 7.9 years (mean of 3.66 year, SD = 2.08)</p> <p>Interrater agreement. 45 assessments were</p>	<p>Males only</p>

<p>Worling and Langton (2015)</p>	<p>SD=1.53)</p> <p>10/191 in custody settings 181/191 in community settings</p> <p>All had sexually offended. Most against younger children (4 or more years younger). Most had offended sexually against multiple victims. Almost a third had received ;prior sanctions from an adult for a previous sexual crime.</p>		<p>and interviews with adult respondents such as parents and/or caregivers, child welfare officers, treatment providers or probation officers. ERASOR ratings were made by clinicians immediately following comprehensive assessments and all clinicians received at least 1 day of training from the first author regarding the development and use of the ERASOR. Mean number of ERASOR ratings completed by the clinicians in this investigations was 7.70 (SD=14.95)</p> <p>High risk: one SD above the mean</p> <p>Below 50th percentile – differentiated low from moderate risk.</p> <p>Low score: 0-25 Moderate score: 26-33 High score: ≥34</p>	<p>available from pairs of clinicians who worked jointly. Intraclass correlation coefficients (ICC) were calculated using a one-way random effects model with absolute agreement. Measures for the total score, the clinical judgement rating and the individual risk factors – significant interrater agreement for the all three.</p> <p>Recidivism Data from three sources: national database, provincial database of criminal charges and a provincial database of adult criminal charges.</p> <p>18/191 (9.4% charged with sexual re-offence during the follow-up period.</p> <p>Clinical judgment ratings (high, moderate or low) from the ERASOR were not</p>	
-----------------------------------	---	--	--	--	--

				<p>significantly related to sexual recidivism over the F-U period, AUC = 0.61 (95% CI 0.48 to 0.74) p=0.13</p> <p>Total score was significantly predictive of sexual reoffending AUC=0.72 (95%CI 0.61 to 0.83), p=0.002</p> <p>Simple sum of risk factors rated as present: AUC= 0.73 (95% CI 0.63 to 0.84) p=0.001</p> <p>ERASOR total score and the sum of risk factors rated present were not significantly different z=0.50, p=0.615</p>	
				<p>Use of cut-off score was significantly predictive of sexual reoffending: AUC=0.65 (95% CI 0.51 to 0.80) p=0.04</p> <p>Short term risk – tested on 70 adolescents who were followed up for between 0.1 and 2.5</p>	

				<p>years (mean=1.4, SD =0.71). The sexual recidivism rate of 8.6% 6/70 for those in the shorter tem F-U period was comparable to that found with the full sample. Using this briefer time frame, clinical judgment ratings were significantly predictive of recidivism: AUC=0.82 (95% CI 0.69 to 96, as was both the total score: AUC = 0.93 (95% CI 0.86 to 0.99) And the sum of risk factors rated as present, Z=0.58, p=0.550</p> <p>Nonsexual Recidivism 13.6% 26/191 of the adolescents had charges for nonsexual violent crimes recorded on one or both of the two official databases.</p> <p>ERASOR total score AUC =0.65 (95% CI 0.53 to 0.76) p=0.017</p>	
--	--	--	--	--	--

				<p>Simple sum of risk factors rated as present: AUC=0.64 (95% CI 0.52 to 0.76) p=0.026</p> <p>Clinical judgment rating AUC=0.61 (95% CI (0.50 to 0.73) p = 0.468 and were unrelated to nonviolent recidivism for this time frame.</p> <p>Nonviolent reoffending 35/191 18.32% were charged with new nonviolent crimes during the total follow-up interval.</p> <p>ERASOR total score, AUC=0.57 (95% CI 0.47 to 0.68) p=0.176</p> <p>Simple sum of risk factors rated present AUC=0.65 (95% CI (0.41 to 0.90) p=0.177</p> <p>Clinical judgment ratings, AUC=0.61 (95% CI 0.41 to 0.82) p=0.328 and were not predictive of</p>	
--	--	--	--	--	--

				subsequent nonviolent crimes.	
<p>Worling & Langton (2015)</p> <p>To identify protective factors that are predictive of desistance from reoffending who have offended sexually.</p>	<p>81 adolescent males (subset of above study)</p> <p>Aged between 12-19 years at the time of assessment. Mean = 15 (SD 1.53)</p> <p>Convicted of and/or acknowledged criminal sexual behaviours.</p>		<p>ERASOR and BERS-2-2</p> <p>Testing for ERASOR as above. At the time of initial assessment parents or caregivers were asked to complete the BERS-2 as part of a comprehensive assessment of strengths and risks.</p> <p>The BERS-2 –Parent Rating Scale was used to collect information from caregivers regarding the youth’s strengths.</p>	<p>Sexual recidivism. Data obtained as above. F-U period of this study ranged from 0.1 to 7.9 years. (mean 3.66, SD 2.08)</p> <p>Sexual recidivism rate was 7/81 (8.6%)</p> <p>Non sexual recidivism was 9/81 (11.1%)</p> <p>BERS-2 Affective Strength scale was found to be significantly predictive of desistance from sexual reoffending.</p> <p>See table of outcomes for data.</p>	<p>Small sample size as is the absolute number os sezual and nonsexual recidivists. Also no precise dates of sexual recidivism to determine the exact at risk periods. Most of the adolescents whow ere the focus of the study were subsequently involved in specialized treatment focused on reducing their risk of reoffend. It is likely that the actual re-offence rate is higher than that based on official data and this is particularly true for sexual crimes as sexual crimes are rarely disclosed to authorities</p>
<p>Elkovitch et al (2008)</p> <p>To examine the 1) accuracy of clinical judgments of risk made after completing risk assessment instruments</p>	<p>166 adolescent males Referred to a residential sex offender program.</p> <p>Inclusion criteria</p>		<p>SAVRY J-SOAP II</p> <p>Graduate student raters assessed risk of reoffending for each participant using two risk</p>	<p>Follow up mean = 80.01 months, SD= 42 months. 8.4% (n=14) youth were charged with sexual felonies,</p>	<p>The ROC AUCs indicated a near-zero relationship between risk classification and violent recidivism, including sexual recidivism.</p>

<p>2) whether instrument informed clinical judgments made with a high degree of confidence are associated with greater accuracy</p> <p>3) the risk assessment instruments and subscales most predictive of clinical judgments.</p> <p>USA</p>	<p>Aged between 13-17 years</p> <p>Intellectual and adaptive functioning at least at the borderline level</p> <p>Committed a sexual offence and mandated to receive treatment</p> <p>Demonstrated self-control that would allow functioning in an open, unlocked treatment program.</p> <p>Period of at least 250 days must have elapsed since program completion so that youth could be followed to assess reoffending.</p> <p>Mean age at admission: 15.31 years (SD=1.51)</p> <p>Majority of youth's index offences were perpetrated against youth who were at least 3 years younger than themselves (79.5%, n=132)</p>		<p>assessment instruments. Ratings completed based on reviews of comprehensive psychological file information which included psychological evaluations and on-going psychological assessments, psychiatric reports, medical and psychopathological evaluations, social work reports, treatment plans, therapy progress notes, teacher assessments and school records, criminal information and legal reports.</p> <p>Rates given training. Raters made a clinical judgment as to whether each youth was at low, moderate, or high risk of engaging in post release sexual violence and nonsexual violence. Raters were instructed that, while they could use the SAVRY and J-SOAP-II to guide their judgments, they could place as little or as much weight on the instruments as they wanted.</p>	<p>10.2% (n=17) were charged with nonsexual violent offences, and 20.5% (n=34) had been charged with any violent offence.</p> <p>Recidivism</p> <p>Juvenile and adult arrest and legal processing records for each youth were obtained from state law enforcement sources. To take into account the possibility of dispositional bargaining, charges, rather than convictions were measured. Also measured through unofficial means, including treatment records.</p> <p>Inter-Rater Reliability</p> <p>A random sample of 21.9% (n=37) intraclass correlation coefficients for single raters were calculated. The ICC_{1s} for</p>	<p>Raters were no more accurate than chance in predicting either sexual recidivism or nonsexual violent recidivism.</p>
---	--	--	--	---	---

				instrument informed clinical judgments of nonsexual violence and sexual violence fell in the 'excellent' range. ICCs for nonsexual violence and sexual violence were 0.78 and 0.79 respectively. J-SOAP II and SAVRY total scores were found to be highly correlated with instrument informed clinical judgments.	
Epperson and Ralston (2009) USA Data from a slide presentation	Utah Study 494 males, Juvenile sex offenders between the ages of 11 and 16.99 years.		JSORRAT-II 12 variable that comprise the JSORRAT-II	Follow up period 24 months. ICC for student coders 0.68 Scored the same 16 cases Reliability of evaluators 0.91 Scored the same 17 cases. Recidivism Total 69/494 (14%)	
Viljoen et al (2007) Canada	N=169 Male adolescents admitted to a nonsecure residential	All raters were blind to youths' subsequent charges and	Trained raters completed the J-SOAP-II, SORRAT-II and SAVRY for each youth.	Follow-Up period Minimum 250 days post discharge.	

<p>The present study examined the ability of the J-SOAP-II, SAVRY and J-SORRAT-II to predict whether male adolescents in a residential sex-offender treatment programme engaged in sexual and nonsexual violence during and following treatment.</p>	<p>treatment program for sexually abusive adolescents in a city in the US.</p> <p>The adolescents must be adjudicated by the courts for sexual or nonsexual offences and ordered to receive treatment.</p> <p>This sample included all youth admitted to the program from 1992 to 2005.</p> <p>Mean age at time of admission: 15.37 (SD 1.51) Caucasian 141/169 (83.4%) African American 8/169 (4.7%) Hispanic 8/169 (4.7%) American Indian or Alaskan Native 2/169 (1.2%) Mixed race 4/169 (2.4%)</p> <p>Majority of index offences were perpetrated against victims who were at least 3 years younger than themselves (140/169 82.8%)</p>	<p>convictions.</p>	<p>Files had following components; psychiatric assessments, psychological assessments, nursing records, medical examination information, social work reports, teacher assessments, school records, treatment plans, progress notes, physician orders and correspondence with court, arrest records and other treatment providers.</p>	<p>Followed for an average of 6.58 years following discharge from the treatment program (SD 3.49, range = 280 days to 12.01 years)</p> <p>Recidivism Information gathered both during and following the treatment program by examining law enforcement, probation and treatment records.</p> <p>On average youth spent approx. 1 year in the treatment period (m=389.7 days, SD 232.3) during which time their aggressive behaviours were examined.</p> <p>ICC 'sufficient'</p> <p>To assess sexual and nonsexual violence after discharge, youth's statewide juvenile justice and adult criminal records were obtained through law enforcement and</p>	
--	--	---------------------	---	---	--

				probation sources as well as by reviewing subsequent treatment records. A rater, who was not involved in completing the risk assessment tools, coded whether the youth was charged with sexual offences, any nonsexual violent offences, serious nonsexual violent offences, and any offences following discharge.	
Viljoen et al (2009)	Caucasian 160/193 (82.9%) African American 16/193 (8.3%) Hispanic 9/163 (1.6%) American Indian or Alaskan Native 3/193(1.6%) Unknown 5/193 (2.6%)		Data was collected by two trained doctoral students in clinical forensic psychology. Raters were blind to youth's subsequent charges and convictions		
White et al (2008) USA To investigate the clinical utility of a measure of CU traits, the Inventory of Callous-Unemotional Traits	172 boys detained in a secure custody facility in the US. The sample are consecutive admissions of boys who were court ordered into secure custody following disposition for a sexual offence during a 41 month				

<p>(ICU) by testing patterns of convergent validity across both self-report and parent-report versions of this instrument with two risk assessment measures commonly used to assess sexual and nonsexual risk factors among sexually offending youth.</p> <p>Retrospective cohort study</p>	<p>time period. Youth with missing data >20% was excluded. This left a final sample of 94 boys Age range from 12-18 years (mean 15.22, SD 1.48).</p> <p>Approximately equal numbers of African-American (45.7%), and non-Hispanic White male youth (53.2%) with a very small number of Hispanic and biracial youth (1.1%)</p>				
---	--	--	--	--	--

