

Catalyst Consortium
Understanding and Assessment
**Australian Risk Assessment Tools
Evaluation Directory**

January 2020



Centre for Forensic
Behavioural Science



catalyst
CONSORTIUM

This report was prepared as part of the Catalyst Consortium program of work which is funded by the Department of Justice and Regulation, and the Victorian Institute of Forensic Mental Health (Forensicare).

The structure of the report is based on the Risk Assessment Tools Evaluation Directory developed by the Scottish Risk Management Authority (RATED). The content of the Australian version of the RATED (Aus-RATED) is not endorsed by the Scottish Risk Management Authority and the views of the authors do not necessarily represent the views of Forensicare, the Government of Victoria or the Scottish Risk Management Authority.

Suggested Citation:

Catalyst Consortium. (2020). Australian Risk Assessment Tools Evaluation Directory. Melbourne, Australia: Catalyst Consortium.

Acknowledgments:

The work in this document is based on the Risk Assessment Tool Evaluation Directory developed by the Scottish Risk Management Authority (RATED).

Contact:

Centre for Forensic Behavioural Science
Level 1, 582 Heidelberg Road
Alphington Victoria 3078
Australia
Info-CFBS@swin.edu.au
+61 3 9214 3887

Contents

Introduction	1
Identification of studies and inclusion criteria	2
Using Aus-RATED to select a risk assessment tool	3
Criteria definitions and performance management scales	5
Empirical grounding	5
Inter-rater reliability	5
Validation history	6
Contribution to risk practice	6
Other considerations	6
Statistical Terminology	7
Risk Assessment Tools Summary Table	8
General Risk Assessment	
Level of Service Inventory-Revised	9
Level of Service Inventory-Revised: Screening Version	14
Level of Service/Case Management Inventory and Level of Service/Risk Need Responsivity	18
Violence Risk Assessment	
Historical Clinical Risk-20 Version 2	24
Historical Clinical Risk-20 Version 3	32
Violence Risk Scale	37
References	41
Level of Service Inventory-Revised	41
Level of Service Inventory-Revised: Screening Version	42
Level of Service/Case Management Inventory and Level of Service/Risk Need Responsivity	42
Historical Clinical Risk-20 Version 2	43
Historical Clinical Risk-20 Version 3	44
Violence Risk Scale	45



Introduction

The Risk Assessment Tools Evaluation Directory (RATED) (www.rma.scot/research/rated/) was originally developed by the Risk Management Authority in Scotland in 2006 (with the fourth edition published in September 2019) and constitutes a summary of risk assessment tools and the empirical evidence underpinning their utility across a number of offending domains.

The RATED is intended to assist practitioners to apply appropriate risk assessment tools as part of a structured approach to assessment, in which risk assessment instruments can assist with the identification of risk factors, needs and strengths of an individual. It provides relevant research information on each instrument included in the Directory and highlights the strengths and limitations that the assessor should take into account when applying a tool as part of a holistic risk assessment process. It aims to provide a summary of the empirical evidence to inform a balanced individualised approach to assessment and to contribute to effective and ethical practice.

The RATED was originally developed with reference to risk assessment validation studies most relevant to the UK population and in support of the Framework for Risk Assessment Management and Evaluation (FRAME) (www.rma.scot/standards-guidelines/frame/). This framework promotes risk assessment practice that makes meaningful use of risk assessment tools without being overly reliant on them, ensuring that the valuable contribution of such instruments is located within a structured approach which recognises the strengths and limitations of tools and the importance of professional/clinical judgment for making meaningful clinical decisions with risk assessment information derived from the tools.

The Catalyst Consortium has now partnered with the Risk Management Authority to produce an adapted version of the RATED tailored to the Australian setting (Aus-RATED). The present report focuses on risk assessment and management tools for general and violent offending among adults that are widely used in Australian jurisdictions. Future editions of the Aus-RATED will be expanded to include an examination of assessment tools specific to sexual violence risk and family violence risk.

Identification of studies and inclusion criteria

The Aus-RATED uses a standard pro-forma developed by the Risk Management Authority to describe each tool based on the summarised evidence and the adopted evaluation framework. The pro-forma provides individual assessments of the tool against framework criteria which are considered essential for the evaluation of risk assessment tools, as adopted by the Risk Management Authority (2019). Research evidence collated for the Aus-RATED is drawn from published studies identified via academic research databases. Applying the approach developed by the Risk Management Authority, publications were assessed on the basis of the following criteria:

- **Authorship**

whether the author(s) were, or were not, involved in the validation of the tool under consideration with greater weight given to the studies conducted by independent researchers other than the authors of the risk assessment tool because this better reflect the tool's real-world validity;

- **Place of publication**

with greater weight given to publications in peer-reviewed journals;

- **Date of publication**

with greater weight given to more recent publications;

- **Sponsorship**

whether the study has been commissioned and/or funded by a government or statutory authority with greater weight given to independent studies;

- **Relevance to the offender population**

whether the study focused on offender populations;

- **Size of study population**

with greater emphasis on evidence drawn from large sample populations;

- **Focus of study**

taking into account whether specific issues have been considered, such as predictive validity, inter-rater reliability, significance and specificity; and

- **Rigorous statistical analyses**

with greater weight given to studies that have used ROC/AUC analyses to assess predictive validity (to enable more informed comparisons between studies).

A note on language and terminology

The Aus-RATED endeavours to use inclusive and respectful language throughout. However, we recognise that any single label or phrase inherently simplifies meaningfully varied experiences across diverse groups of people, and acknowledge the complexity of experiences that sit within the broad labels we have chosen. In addition, language and culture change over time and, while the Aus-RATED aims to reflect recommended practice at the time of writing, it may include content that is in contention or under debate. We welcome feedback and will revise the Aus-RATED over time as appropriate.

Using Aus-RATED to select a risk assessment tool

When using Aus-RATED, practitioners should consider the following:

- **The type of risk to be evaluated (e.g. general, violence, sexual);**
- **The age, gender, race and ethnicity, mental state (at the time of the assessment) and cognitive abilities of the offender;**
- **The applicability of a tool to a particular offending population or minority group; and**
- **The performance of tools with respect to the criteria outlined in the Aus-RATED which includes their validation history, empirical grounding, inter-rater reliability and ability to identify targets for intervention.**

Where practitioners have the necessary competences and training, they can use with confidence tools that possess a robust validation history and empirical grounding. These tend to be tools that have also evidenced high inter-rater consistency, specificity and sensitivity in identifying individuals at risk of reoffending. Sometimes, some of these qualities are present, but not others. Practitioners should be cautious when using tools that possess some but not all of the essential attributes, but may have the potential to demonstrate strong psychometric properties with further studies and/or evidence. Reports detailing findings based on the use of such tools should outline the limitations of their use in this respect and provide a justification for their selection. If an alternate tool, with established empirical support is available then this other tool should be used.

If there is a need for an assessment but no tool with strong psychometric properties is available, the practitioner should consider using the items from an existing tool as a guide, focussing not on the presence of each risk factor but rather on the relevance of each risk factor for the particular individual, while avoiding confident statements about predicted likelihood of recidivism. Instead, practitioners should focus their discussion on the known base rate of the outcome of concern, if this is known, as the base rate is the most accurate estimate of likelihood when little else is known about specific risk factors related to the outcome in the specific sub-population.

Practitioners should give careful consideration to whether the selection of a given instrument is suitable for a particular offending population or minority group (e.g. female offenders, mentally disordered offenders, offenders with an intellectual disability, Indigenous offenders and offenders from culturally and linguistically diverse backgrounds). A good guide to an instrument's suitability for a particular individual is whether a sample from the target group has been included in the original research sample or been the subject of a subsequent successful pilot study.

In all cases, practitioners should have the necessary training for using the tool(s) of their choice, be aware of a tool's limitations and the caveats of its use, and be in a position to discuss these limitations and to evidence their assessment. Importantly, these limitations should be clearly outlined in the assessor's report to inform future decision-making. This is of particular importance when decisions are being made by those who are unfamiliar with risk assessment instruments and their inherent limitations.

When communicating risk assessment information using empirically valid instruments, practitioners should limit conclusions to those supported by existing empirical evidence. Currently, methodologies in risk assessment research support statements about risk discrimination (e.g. "Offenders with a score of 8 are two times more likely to recidivate than offenders with a score of 5") but very rarely support statements about risk calibration (e.g. "A group of offenders with a score of 8 are expected to recidivate at a rate of 35.4% over a 2-year period"). As a result, practitioners can be more confident about recommending more rehabilitation and management services for offenders who scored higher than other offenders, but should exercise strong caution about applying specific absolute recidivism rates (for example, "35.4%") to scores derived from risk instruments. To obtain risk calibration information, each Australian jurisdiction must conduct its own large-scale research to determine estimated recidivism rates specific to their jurisdiction. This is because research shows that the recidivism rates associated with scores on the same risk instrument will vary across different jurisdictions.

Using Aus-RATED to select a risk assessment tool

Finally, it is important to note that, whilst effort has been made to ensure that the accuracy of the information presented in the Aus-RATED, the evolving nature of research relating to risk assessment and risk management means that the evidence base is continually subject to change. Practitioners are, therefore, encouraged to keep abreast of the emerging evidence when administering the assessment tools.

Criteria definitions and performance management scales

The following criteria, as adopted by the Risk Management Authority, have been used to identify the strengths and weaknesses of each risk assessment tool in the Aus-RATED.

Empirical grounding

Empirical grounding examines the scientific and theoretical underpinnings of the risk assessment tools. For example, a risk assessment tool based on sound theoretical evidence and/or other extensive scientific findings observed in prior research would be considered to have a high level of empirical grounding. Therefore, higher levels of empirical grounding may increase the utility of the instrument in assessing the risk posed by an individual.

The risk assessment tools are rated on this criterion using the following performance scale:

0 points: No evidence of empirical grounding.



3 points: Limited evidence of empirical grounding.



6 points: Moderate evidence of empirical grounding, but more evidence is required.



9 points: Sufficient evidence of empirical grounding. Tool founded on strong empirical research.



Inter-rater reliability

Inter-rater reliability refers to the degree to which two or more assessors are consistent in their ratings of the risk presented by the individual being assessed using the same risk assessment tool. Thus, it is desirable for a risk assessment tool to have high inter-rater reliability whereby assessors score the items similarly when using the same tools. Researchers must be able to demonstrate that the risk assessment instruments are reliable, since without reliability results, the tool's performance cannot be replicated and validation cannot be attained.

In the literature, inter-rater reliability is the estimation based on the correlation of scores among two or more raters who rate the same scale, item or instrument. Studies summarised in the Aus-RATED have generally used the Intra-Class Correlation Coefficient (ICC) to measure inter-rater reliability.

Tools are rated on this criterion using the following performance scale:

0 points: No evidence of, or evidence indicating poor inter-rater reliability.



3 points: Limited evidence of inter-rater reliability.



6 points: Moderate evidence to suggest that inter-rater reliability exists but insufficient to find this factor fully present for this tool.



9 points: Sufficient evidence of high inter-rater reliability.



Criteria definitions and performance management scales

Validation history

To rate a tool in this area, the existence and quality of validation studies is considered and assessed on the basis of the availability of two or more papers written by different authors in peer reviewed journals. The papers are required to have examined the predictive validity of the tool and/or its practical usefulness for the assessment and management of risk of harm to others. This approach accommodates concerns that have been raised in the literature that different research designs may be appropriate for identifying the properties, strengths and limitations of various types of instruments.

The validation history criterion is split into five subsections:

- **General predictive validity**

The capability of the risk assessment tool to discern the difference in the risk of reoffending between the recidivist and non-recidivist populations.

- **Applicability: Female offenders**

The validity of the risk assessment tool for female offender populations.

- **Applicability: Indigenous offenders and CALD offenders**

The validity of the risk assessment tool for Indigenous offenders (including Aboriginal and Torres Strait Islander offenders and Indigenous offenders in overseas jurisdictions) and offenders from culturally and linguistically diverse (CALD) backgrounds.

- **Applicability: Mentally disordered offenders**

The validity of the risk assessment tool for offenders diagnosed with a mental illness.

- **Applicability: Offenders with intellectual disability**

The validity of the risk assessment tool for offenders with a cognitive impairment or intellectual disability.

Tools are rated against each validation domain using the following performance scale:

0 points: No validation evidence available.



3 points: Limited validation evidence available or available research provides a mixed interpretation of performance of the measure of this criterion.



6 points: Moderate validation evidence available – at least two independent studies in a peer reviewed journal.



9 points: Sufficient validation evidence available – at least three independent studies by different authors in peer reviewed journals.



Contribution to risk practice

This section provides a qualitative evaluation of the properties of the instrument and its contribution to risk practice.

Other considerations

This section includes any other considerations regarding the instrument that may be relevant.

Statistical terminology

The following table summarises several statistics which are used to support the validation of a risk assessment in accordance with the criteria used to rate the risk assessment tools within the Aus-RATED.

	Definition	Interpretation
Inter-Rater Reliability		
Intra-Class Correlation Coefficient (ICC)	An ICC score represents the average correlation between two scores provided by two randomly selected scorers.	ICC values range from 0 to 1. A high ICC close to 1 indicates high similarity between two scores provided by two randomly selected scores, while a low ICC close to zero means that the two scores are not similar. Cicchetti (1994) recommends the following thresholds: <ul style="list-style-type: none"> • .75 to 1.0 = excellent • .40 and .75 = moderate • <.40 = poor These values have also been endorsed by Fleiss, Levin and Paik (2003).
Predictive Validity		
Area Under the Curve (AUC)	AUC values represent the probability that a randomly selected individual who is positive for a particular outcome (e.g., recidivism, parole violation) will score higher on the predictor variable than an individual who does not have that outcome. AUCs are the most commonly used and recommended statistic for risk assessment scales (Helmus & Babchishin, 2017; Mossman, 2013; Rice & Harris, 2005).	AUC values can be interpreted as the probability that a randomly selected recidivist would have a higher score than a randomly selected non-recidivist. The AUC can vary between 0 and 1, with .50 indicating the level of prediction that would be expected by chance, AUCs above .50 demonstrating positive predictive accuracy (ie. higher scores on a measure are associated with a higher level of recidivism), and AUCs below .50 demonstrating negative predictive accuracy (ie. higher scores are associated with a lower likelihood of recidivism). Using Cohen's <i>d</i> values as a guide, Rice and Harris (2005) recommend the following cut-points for AUC interpretation in the social sciences: <ul style="list-style-type: none"> • Low = .56 – .64 • Med = .64 – .71 • High = .71+
Effect size	The effect size quantifies the size of the difference between two groups (e.g. recidivists and non-recidivists) on a variable.	In the Aus-RATED, any statistic that quantifies the magnitude of the relationships between risk scores and outcome can be considered to be an effect size. This includes Cohen's <i>d</i> which, for risk assessment scales, describes how much recidivists differ in risk scores from non-recidivists. A <i>d</i> of 0.5 means that recidivists score half a standard deviation higher than non-recidivists on the risk scale. Roughly, Cohen's <i>d</i> values of .20, .50, and .80 are generally considered small, moderate and large (Cohen, 1998).
Point-Biserial Correlation	Point-Biserial Correlations (<i>r</i>) measure the association between a predictor variable and the outcome (e.g. recidivism).	The values of <i>r</i> can range from -1 to 1, with 0 indicating that there is no relationship between the predictor variable and the outcome. Positive values indicate that high scores are associated with increased recidivism, whereas negative values indicate that high scores are associated with decreased recidivism. It should be noted that correlations can be problematic when applied to data with a dichotomous outcome (e.g. recidivism), as the size of the correlation can be unduly influenced by the base rate (Helmus & Babchishin, 2017). In risk assessment research, the base rate of violence is typically lower than 50% (Rice & Harris, 1995). The further the base rate deviates from 50%, the smaller the correlation becomes, regardless of the true underlying relationship between the variables. This also means that there are no clear cut-points for interpretation.

Risk Assessment Tools Summary Table

	Validation Evidence					
	Age range	General Offender Populations	Female Offenders	Indigenous Offenders & CALD Offenders	Mentally Disordered Offenders	Offenders with Intellectual Disability
General Risk Assessment						
Level of Service Inventory – Revised (LSI-R) Assessor Qualifications: (a) assessors must possess advanced training, certification and experience in psychological assessment or a related discipline; or (b) pass a training course certified by the publishers. Can be used by a large range of professionals including social workers and probation officers.	16+	Limited	Limited	Limited	Limited	None
Level of Service Inventory – Revised: Screening Version (LSI-R:SV) Assessor Qualifications: Similar specifications as that applying to the LSI-R.	16+	Limited	Limited	None	Limited	None
Level of Service/Case Management Inventory (LS/CMI) and Level of Service/Risk Need Responsivity (LS/RNR) Assessor Qualifications: Similar specifications as that applying to the LSI-R.	16+	Moderate	Moderate	Moderate	Moderate	None
Violence Risk Assessment						
Historical Clinical Risk-20 Version 2 (HCR-20^{v2}) Assessor Qualifications: Assessors must possess a degree, certificate or licence to practice in a health care profession or occupation. Assessors must also possess the necessary training and experience in the ethical administration, scoring and interpretation of clinical behavioural assessment instruments.	18–65	Sufficient	Limited	Limited	Moderate	Moderate
Historical Clinical Risk-20 Version 3 (HCR-20^{v3}) Assessor Qualifications: Similar specifications as that applying to the HCR-20 ^{v2} .	18–65	Limited	Limited	None	Moderate	None
Violence Risk Scale (VRS) Assessor Qualifications: No professional qualifications required. Can be used by workers within the criminal justice system. Assessors are required to undertake a training course.	18+	Moderate	None	None	Limited	None

General Risk Assessment

Level of Service Inventory-Revised (LSI-R)

Andrews and Bonta (1995)

Description

- The LSI-R is a 54-item tool designed to assess criminogenic risk and need in general populations of offenders.
- Items are subdivided across ten subsections.
- Generates a composite score of risk-need of the offender. The risk is categorised as either 'minimum', 'medium' or 'maximum'.
- Normed on North American prison, parole and probation populations.

Strengths

- Ability to discriminate risk across various outcome measures such as spousal abuse recidivism (Hendricks et al, 2006).
- Provides structured professional decision-making in a way that is comprehensive and consistent regardless of the case presented (Campbell, French & Gendreau, 2009).

Empirical Grounding

Summary



The LSI-R is founded on strong empirical research regarding criminal behaviour.

- The LSI-R is supported by and reflective of three primary sources of information: (1) prior literature on recidivism; (2) professional opinions of probation officers; and (3) social learning theory of criminal behaviour (Andrews & Bonta, 1995).
- The subscales reflect the main risk factors identified in the research literature (Andrews & Bonta, 2010).

Inter-Rater Reliability

Summary



While there is no Australian research examining the inter-rater reliability of the LSI-R, international research provides good evidence of moderate to high inter-rater reliability, with some variation across the different subscales.

Australian Research

No empirical evidence available.

International Research

- Dahle (2006)
 - Male prisoners, Germany
 - High inter-rater reliability (ICC = .93).
- Hollin, Palmer and Clarke (2003)
 - Male offenders, England
 - 90% agreement rate.
- Lowenkamp et al. (2004)
 - Offender vignettes presented to 167 correctional practitioners trained in use of LSI-R, US
 - Moderate to high levels of agreement observed across subscales ranging from 61.5% to 97.7%.
- Palmer and Hollin (2007)
 - Female offenders, England
 - Inter-rater agreement levels of 95%.
- Persson et al. (2017)
 - Forensic psychiatric patients, Sweden
 - High inter-rater reliability (ICC = .92).

Level of Service Inventory-Revised (LSI-R)

General Predictive Validity

Summary



Research, in both Australia and internationally, provides mixed evidence for the predictive accuracy of the LSI-R in general offending populations, with findings ranging from low to high predictive validity across different jurisdictions.

Australian Research

- Hsu, Caputi and Byrne (2009)
 - 27,822 offenders (21,916 male, 5906 female), NSW, reoffending, mean follow-up period eight monthsScores had strongest relationship with recidivism for male prisoners ($r = .20$), female prisoners ($r = .23$) and female community-based offenders ($r = .21$). Increased total score was associated with a greater likelihood of reoffending by a factor of 1.03 for male offenders and 1.05 for female offenders. Criminal History was the strongest subscale across gender and sentence orders.
- Watkins (2011)
 - 7555 male prisoners, NSW, reincarceration, two year follow-up periodModerate predictive accuracy (AUC = .69).

International Research

- Duwe and Rocque (2016)
 - 26,001 male and female prisoners, US, reconviction in three years post-releaseLow predictive accuracy (AUC = .63).
- Georgiou (2019)
 - 10,000 offenders (1619 prisoners; 8381 community-based sentence) (7751 male; 2249 female), US, reconviction within 36 months after becoming at-risk in the communityModerate predictive accuracy for general (AUC = .68) and violent recidivism (AUC = .64).
- Hausam, Lehmann and Dahle (2018)
 - 272 male prisoners, Germany, institutional misconduct (mean follow-up period 17 months) and reoffending (mean follow-up period 22 months post-release)Low predictive accuracy for violent and non-violent institutional misconduct (AUC = .63 and .64 respectively). Moderate predictive accuracy for non-violent reoffending (AUC = .66). High predictive accuracy for violent recidivism (AUC = .71)
- Lowenkamp, Lovins and Latessa (2009)
 - 483 offenders on probation (369 male, 116 female), US, rearrest and reincarceration, 1.5 year follow-up periodCorrelation between total score and rearrest was $r = .36$, and $r = .33$ for total score and reincarceration.
- Manchak, Skeem and Douglas (2008)
 - 1144 male prisoners convicted of serious violent offence, US, reoffending, mean follow-up period 28 months post-releaseHigh predictive accuracy (AUC = .73) for both general and violent recidivism.
- Ostermann and Herrschaft (2013)
 - 900 parolees (450 male, 450 female), US, rearrest, reconviction and parole violation, three year follow-up periodLow predictive accuracy for rearrest (AUC = .63), reconviction (AUC = .62) and parole violation (AUC = .62).
- Zhang (2016)
 - 112 male offenders, China, reoffending, mean follow-up period two yearsHigh predictive accuracy (AUC = .73).

Level of Service Inventory-Revised (LSI-R)

Applicability: Female Offenders

Summary



Studies indicate that the LSI-R performs similarly for both male and female offenders, with mixed evidence for the predictive validity of the LSI-R ranging from low to high predictive validity in samples of female offenders.

Australian Research

- Watkins (2011)
 - 614 female prisoners, NSW, reincarceration, two year follow-up period
 - Moderate predictive accuracy (AUC = .69).

International Research

- Manchak et al. (2009)
 - 56 female prisoners convicted of serious violent offence, US reoffending in 12 months post-release
 - High predictive accuracy (AUC = .77).
- Ostermann and Herrschaft (2013)
 - 450 female parolees, US, rearrest, reconviction and parole violation, three year follow-up period
 - Low predictive accuracy for rearrest (AUC = .62), reconviction (AUC = .62) and parole violation (AUC = .62).
- Ostermann and Salerno (2016)
 - 4,727 female community-based offenders, US, rearrest or parole revocation in one year post-release
 - Low predictive accuracy (AUC = .624).
- Smith, Cullen and Latessa (2009)
 - Meta-analysis of 25 studies
 - Mean correlation (r) of .34 with recidivism (random effects, 95% confidence interval = .29 to .41, $k = 27$). Shorter follow-up periods (12 months or less) were associated with larger correlations ($r = .43$, 95% confidence interval = .41 to .45, $k = 27$) than longer time intervals ($r = .28$, 95% confidence interval = .26 to .30, $k = 27$). Concluded that relationship between LSI-R and recidivism for female offenders is statistically and practically similar to that for males.

Level of Service Inventory-Revised (LSI-R)

Applicability: Indigenous Offenders & CALD Offenders

Summary



International evidence for the applicability of the LSI-R to Indigenous offenders and CALD offenders is mixed, with some subscales (such as criminal history and alcohol/drug) appearing to be less predictive for these offender groups. Research regarding its applicability to Aboriginal and Torres Strait Islander offenders, albeit limited, provides tentative support for the use of the LSI-R with this offender group, although one large study indicates low predictive accuracy.

Australian Research

- Hsu, Caputi and Byrne (2010)
 - 13,911 Aboriginal and Torres Strait Islander offenders (10,958 male, 2,953 female), NSW, reoffending, mean follow-up period eight months
 - Correlation between total score and recidivism was $r = .12$ for males and $r = .16$ for females. Correlations for Aboriginal and Torres Strait Islander offenders were smaller than correlations for non-Aboriginal and Torres Strait Islander offenders (male $r = .18$; female $r = .21$). Increased total score was associated with a greater likelihood of reoffending by a factor of 1.03 for male Aboriginal and Torres Strait Islander offenders and 1.05 for female Aboriginal and Torres Strait Islander offenders. Criminal History was the strongest subscale for both Aboriginal and Torres Strait Islander and non-Aboriginal and Torres Strait Islander offenders.
- Watkins (2011)
 - 2,465 male Aboriginal and Torres Strait Islander prisoners, NSW, reincarceration, two year follow-up period
 - Moderate predictive accuracy (AUC = .66).

International Research

- Chenane et al. (2015)
 - 2,778 male prisoners, US, institutional misconduct in the first two years of imprisonment
 - Total score and subscale scores similarly predicted the prevalence of institutional misconduct for White, Black and Hispanic/Latino prisoners but showed greater predictive utility for White prisoners than Black or Hispanic/Latino prisoners when predicting the frequency of institutional misconduct.
- Fass et al. (2008)
 - 975 male offenders, US, rearrest in 12 months post-release
 - Inconsistent validity with racial and ethnic minority offender groups. Composite score correctly predicted re-arrest for 80% of White offenders, 82% of Hispanic/Latino offenders and 43% of Black offenders. AUC values for all racial and ethnic groups were low (0.55 for White offenders, 0.61 for Black offenders and 0.54 for Hispanic/Latino offenders).
- Ostermann and Salerno (2016)
 - 9454 community-based offenders (4727 male, 4727 female) (5647 Black, 2455 White, 1352 Hispanic/Latino), US, rearrest or parole revocation in 12 months post-release
 - Low predictive accuracy for Black and White offenders (AUCs = .61 and .62 respectively). Moderate predictive accuracy for Hispanic/Latino offenders (AUC = .66).
- Wilson and Gutierrez (2014)
 - Meta-analysis comprising 10 Canadian studies, one US study and one Australian study, reoffending (general and violent), mean follow-up period for combined sample 29.7 months
 - Total score significantly predicted general recidivism with an effect size of $d = .29$ (random effects, 95% confidence interval = .23 to .36, $k = 3$). All eight subscales significantly predicted general and violent recidivism for Indigenous offenders, and were found to predict violent recidivism similarly for both groups. For general recidivism, predictive accuracy of total score and five subscales (Education/Employment, Companions, Alcohol/Drugs, Procriminal Attitude and Criminal History) was significantly lower for Indigenous compared to non-Indigenous offenders. Concluded that, while the LSI-R predicts recidivism with Indigenous offenders, it does so with less accuracy than with non-Indigenous offenders.

Level of Service Inventory-Revised (LSI-R)

Applicability: Mentally Disordered Offenders

Summary



There is limited international research regarding the validity of the LSI-R amongst mentally disordered offenders, with one study providing good support for its utility in this group. There is no Australian research examining the predictive validity of the LSI-R in Australian mentally disordered offender populations.

Australian Research

No empirical evidence available.

International Research

- Persson et al. (2017)
 - 200 forensic psychiatric patients (174 male, 26 female), Sweden, violent act, 12 month follow-up period
 - High predictive accuracy for total score (AUC = .73).

Applicability: Offenders with Intellectual Disability

Summary



There is no research evidence, in Australia or internationally, regarding the predictive validity of the LSI-R amongst offenders with a cognitive impairment or intellectual disability.

Australian Research

No empirical evidence available.

International Research

No empirical evidence available.

Contribution to Risk Practice

- The LSI-R has the ability to create awareness of a number of static and dynamic risk factors pertinent to the individual's general risk of recidivism. Information obtained through the LSI-R can inform the level and focus of monitoring and supervision strategies.
- The LSI-R can aid ongoing evaluation of an individual's risk of reoffending and their criminogenic needs. In this regard, studies have examined the importance of change on the LSI-R in the prediction of recidivism and found that more recent, proximal LSI-R scores were more predictive of subsequent recidivism than earlier scores (e.g. Labrecque et al., 2014).

Other Considerations

- Fewer validation studies conducted with other populations such as Indigenous offenders, CALD offenders, and mentally disordered offenders.
- Translations into a country's native language have demonstrated the reliability and predictive validity of the LSI-R in Guangzhou, China (Zhang & Liu, 2015).
- Requires refresher training – experience and training in the LSI-R can affect the reliability of the instrument (Lowenkamp, Lovins & Latessa, 2009).
- The tool is a quantitative survey of risk-need factors that are supported by research, professional opinion and social learning theory on criminal behaviour. It is not a comprehensive measure of mitigating and aggravating risk factors related to offender risk practices (Andrews & Bonta, 1995).
- The LSI-R should be completed using information obtained from interviews with the individual and other collateral sources of information.

Level of Service Inventory-Revised: Screening Version (LSI-R:SV)

Andrews and Bonta (1998)

Description

- The LSI-R:SV is an 8 item actuarial screening tool derived from the LSI-R.
- Similar categorisation of risk as observed in the LSI-R. High composite scores may warrant further analysis from the full LSI-R or LS/CMI assessment.
- Normed on Canadian institutionalised and probation populations.

Strengths

- Ideal for use when a complete LSI-R assessment may not be feasible, due to time constraints or insufficient staff resources.
- The LSI-R:SV can assist in prioritising cases for further intervention including assessment.

Empirical Grounding

Summary



The LSI-R:SV is founded on strong empirical research regarding criminal behaviour.

- Developed in part from the LSI-R, a well-validated tool with the developments informed by further research and consultation with practitioners as well as general personality and social learning perspective theories (Andrews, Bonta & Wormith, 2004).
- Two principles guided the creation of the LSI-R:SV: (a) item selected from the LSI-R must have demonstrated the ability to predict recidivism; and (b) the majority of items must be dynamic to remind the user that the instrument is designed to inform treatment decisions.

Inter-Rater Reliability

Summary



Research regarding the inter-rater reliability of the LSI-R:SV is limited but indicates moderate to high inter-rater reliability. There is no Australian research examining the inter-rater reliability of the LSI-R:SV.

Australian Research

No empirical evidence available.

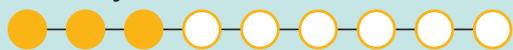
International Research

- Livingston et al. (2015)
 - Offenders on probation, Canada
 - High inter-rater reliability (ICC = .79).
- Walters (2011)
 - Male prisoners, US
 - Good inter-rater reliability (ICC = .71).

Level of Service Inventory-Revised: Screening Version (LSI-R:SV)

General Predictive Validity

Summary



There is no research regarding the predictive validity of the LSI-R:SV in Australian offender populations. International research (predominantly in the US and Canada) indicates low to moderate predictive accuracy amongst general offender populations.

Australian Research

No empirical evidence available.

International Research

- Lowenkamp, Lovins, and Latessa (2009)
 - 483 offenders on probation (369 male, 116 female), US, rearrest and re-incarceration, 1.5 year follow up periodEffective in discerning between low, moderate and high risk offenders. Rearrest rates for low risk offenders was 15%, 34% for moderate risk offenders and 50% for high risk offenders. A similar pattern was found for incarceration rates: low 13%; moderate 29%; high 39%.
- McCafferty and Scherer (2017)
 - 1,900 prisoners (1,581 males and 337 females), US, rearrest in 15 months post-releaseLow predictive accuracy (AUC = .59).
- Walters (2011)
 - 178 male offenders, US, reoffending (general and violent), mean follow-up period 20 monthsDid not predict general (AUC = .57) or violent recidivism (AUC = .55).
- Walters and Schlauch (2008)
 - 159 male prisoners, US, institutional infractions, two year follow-up periodLow predictive accuracy (AUC = .63).
- Yessine and Bonta (2006)
 - 256 high risk male offenders, Canada, reconviction, 3.4 years mean follow-up periodModerate predictive accuracy for general and violent reconviction (AUC = .68 and .67 respectively). Low predictive accuracy for non-violent reconviction only (AUC = .63).

Applicability: Female Offenders

Summary



There is no research regarding the predictive validity of the LSI-R:SV in Australian female offender populations. International research, albeit limited, has found that the LSI-R:SV demonstrates low to moderate predictive accuracy amongst female offenders.

Australian Research

No empirical evidence available.

International Research

- Lowenkamp, Lovins and Latessa (2009)
 - 116 female offenders on probation, US, rearrest and reincarceration, 1.5 year follow-up periodUnable to discriminate across LSI-R risk categories, whereby the confidence intervals for low and moderate risk categories for female rearrests and reincarcerations overlapped. Correlation between total score and rearrest was $r = .22$ (95% confidence interval = .04 to .39), and $r = .19$ for total score and reincarceration (95% confidence interval = .01 to .37).
- McCafferty and Scherer (2017)
 - 337 female prisoners, US, rearrest in 15 months post-releaseLow predictive accuracy (AUC = .62).

Level of Service Inventory-Revised: Screening Version (LSI-R:SV)

Applicability: Indigenous Offenders & CALD Offenders

Summary



There is no research evidence, in Australia or internationally, regarding the predictive validity of the LSI-R:SV amongst Indigenous offenders and CALD offenders.

Australian Research

No empirical evidence available.

International Research

No empirical evidence available.

Applicability: Mentally Disordered Offenders

Summary



Australia research indicates that the LSI-R:SV demonstrates low predictive accuracy for institutional violence and moderate predictive accuracy for general recidivism amongst mentally disordered offenders as a group. There are, however, indications that the LSI-R:SV may not perform as well amongst offenders with a dual diagnosis. There are no international studies examining the applicability of the LSI-R:SV in mentally disordered offender populations.

Australian Research

- Daffern et al. (2005)
 - 232 forensic psychiatric inpatients (193 male, 39 female), Victoria, institutional violence, 12 month follow-up periodLow predictive accuracy for institutional aggression (AUC = .60) and institutional violence (AUC = .59).
- Ferguson, Ogloff and Thomson (2009)
 - 208 psychiatric offenders (157 male, 51 female) released from secure forensic hospital, Victoria, reconviction, follow-up period not reportedRecidivism predicted at a level significantly above chance for any reconviction (AUC = .67), non-violent reconvictions (AUC = .65) and violent reconvictions (AUC = .60). Predictive validity was higher for non-substance abusers (AUC = .78 any reconviction; AUC = .74 non-violent reconvictions; AUC = .71 violent reconviction). For those with a dual diagnoses, the LSI-R:SV performed at levels no better than chance.

International Research

No empirical evidence available.

Applicability: Offenders with Intellectual Disability

Summary



There is no research evidence, in Australia or internationally, regarding the predictive validity of the LSI-R:SV amongst offenders with a cognitive impairment or intellectual disability.

Australian Research

No empirical evidence available.

International Research

No empirical evidence available.

Level of Service Inventory-Revised: Screening Version (LSI-R:SV)

Contribution to Risk Practice

- The LSI-R:SV can aid the assessor in identifying some static and dynamic risk factors pertinent to the individual's likelihood of reoffending.
- The tool is useful for a brief scan of the main risk factors.
- The tool can alert assessors to the need to conduct a more thorough assessment.

Other Considerations

- Some research has found that the LSI-R:SV does not discriminate between moderate and high risk offenders (Lowenkamp, Lovins, & Latessa, 2009).
- The effectiveness of the LSI-R:SV for screening the offending population is based on preliminary and limited evidence (Lowenkamp, Lovins, & Latessa, 2009).
- The LSI-R:SV is "... insensitive to the clinical characteristics and recent hostility that are commonly associated with inpatient aggression ..." (Daffern, 2007, p. 122).
- Assessors should note that this tool is a screening version of the full assessments (i.e. LSI-R, LS/CMI) and is not a comprehensive measure of risk, need and responsivity factors.
- The LSI-R:SV should be completed using information obtained from interviews with the individual and other collateral sources of information.

Level of Service/Case Management Inventory (LS/CMI) and Level of Service/Risk Need Responsivity (LS/RNR)

Andrews, Bonta and Wormith (2004; 2008).

Description

- The LS/CMI and LS/RNR are measures of risk and need factors.
- The LS/CMI and LS/RNR include five assessment sections, three summary sections, and a case management section that are identical to each other. The LS/CMI includes an additional two sections specifically for case management.
- Section 1 (General Risk/Needs Section) consists of 43 items that are grouped into 8 subsections (the Central Eight). This section is used to predict recidivism.
- Other assessment sections include: Specific Risk/Need Factors; Prison Experience; Institutional Factors; Other Client Issues; Social, Health and Mental Health; Special Responsivity Considerations.
- Risk is categorised into five levels: 'very low', 'low', 'moderate', 'high' and 'very high' risk.
- These tools are designed to assist professionals in management and treatment planning with offenders in justice, forensic, correctional, prevention, and related agencies.
- The LS/CMI and LS/RNR are normed on Canadian and USA probation and institutional populations for male and female offenders. Supplementary norms provide for UK and Singaporean offenders.
- It should be noted that much of the research examining the LS/CMI is applicable to the LS/RNR and vice versa.

Strengths

- Combines risk assessment and case management in a single assessment tool.
- It expands the traditional risk/need assessment instrument to a more comprehensive assessment of the offender by including noncriminogenic needs, prison experience and responsivity considerations.
- Assessors are able to identify strengths in the individual and his/her circumstances.
- It allows for a professional override of degree of service provision based on an assessment of offender strengths and specific risk factors that are not captured in Section 1.

Empirical Grounding

Summary



The LS/CMI and LS/RNR are founded on strong empirical research regarding criminal behaviour.

- Developed in part from the LSI-R, a well-validated tool with the developments informed by further research and consultation with practitioners as well as general personality and social learning perspective theories (Andrews, Bonta & Wormith, 2004).
- Section 1 was informed by a re-analysis of LSI-R item data. The manual includes guidelines for deriving a LS/CMI Section 1 score from LSI-R raw data. As reported in the manual (Andrews, Bonta & Wormith, 2004), the correlation between the LSI-R and LS/CMI is .96 (Rowe, 1999).

Level of Service/Case Management Inventory (LS/CMI) and Level of Service/Risk Need Responsivity (LS/RNR)

Inter-Rater Reliability

Summary



While there is no Australian research examining the inter-rater reliability of the LS/CMI and LS/RNR, international studies have found moderate to high inter-rater reliability for both total and subscale scores.

Australian Research

No empirical evidence available.

International Research

- Canales et al. (2014)
 - Community-supervised offenders diagnosed with a mental disorder, Canada
 - High inter-rater reliability for General Risk/Need total score (ICC = .92) and its subscales (ICCs ranged from .71 to .89).
 - Inter-rater reliability for Specific Risk/Need total score and its subscales was also high (ICCs ranged from .74 to .86).
- Rettinger and Andrews (2010)
 - Female offenders, Canada
 - Moderate to high inter-rater reliability estimates (ICCs ranged from .65 for 'financial problems' to .91 for composite General Risk/Need score).

General Predictive Validity

Summary



Research, both in Australia and internationally, generally supports the use of the LS/CMI and LS/RNR in general offender populations, with some indications that the subscales are stronger predictors of general reoffending but less predictive of violent recidivism.

Australian Research

- Gordon, Kelty and Julian (2015)
 - 569 male community-based offenders, Tasmania, reoffending within 12 months of index offence
 - Moderate predictive accuracy (AUC = .66).

International Research

- Guay and Parent (2018)
 - 3,636 offenders (3,334 male, 302 female) sentenced to less than two years, Canada, rearrest, reconviction and parole violation, mean follow-up period of 18.5 months for rearrest, and two years for reconviction and parole violation
 - Moderate predictive accuracy for rearrest (AUC = .70). High predictive accuracy for reconviction (AUC = .71) and parole violations (AUC = .72).
- Jimenez et al. (2018)
 - 19,344 offenders on probation (14,895 male, 4449 female), US, reoffending, 5.5 year follow-up period
 - Correlation between General Risk/Need Score and reoffending was $r = .21$. The likelihood that a probationer would reoffend increased 1.07 times for each 1-score increase in LS/CMI total score. A probationer selected at random with a very high total score was 8.84 times as likely to reoffend as one selected at random with a very low total score.
- Olver, Stockdale, and Wormith (2014)
 - Meta-analysis of 128 studies
 - Total score demonstrated a weighted mean correlation of $r = .29$ with general recidivism (random effects, 95% confidence interval = .27 to .31; $k = 124$) and $r = .23$ with violent recidivism (random effects, 95% confidence interval = .19 to .27; $k = 39$). Subscales tended to predict general recidivism better than violent recidivism (e.g. Antisocial pattern: $r = .31$ general recidivism and .23 violent recidivism), although Criminal Attitudes predicted types of recidivism equally well ($r = .19$ general recidivism and .18 violent recidivism).

Level of Service/Case Management Inventory (LS/CMI) and Level of Service/Risk Need Responsivity (LS/RNR)

Applicability: Female Offenders

Summary



Research suggests that the LS/CMI and LS/RNR perform similarly for both male and female offenders, but that some subscales (particularly substance abuse) may apply differentially according to gender.

Australian Research

- Gordon, Kelty and Julian (2015)
 - 113 female community-based offenders, Tasmania, reoffending within 12 months of index offence
 - Low predictive accuracy (AUC = .58).

International Research

- Andrews et al. (2012)
 - Meta-analysis of five studies
 - General Risk/Need score demonstrated high predictive accuracy for recidivism for both female (AUC = .83) and male offenders (AUC = .75). While all subscales demonstrated moderate to high predictive validity for both male and female offenders (leading the authors to conclude that the subscales are gender-neutral), the substance abuse subscale demonstrated high predictive accuracy for females and accounted for the increased validity of the General Risk/Need score for female offenders.
- Canales et al. (2014)
 - 39 female community-based mentally ill offenders, Canada, reoffending, 4 years mean follow-up period
 - General Risk/Need score demonstrated moderate predictive accuracy for any reoffending (AUC = .67) but low predictive accuracy for violent reoffending (AUC = .62), while the reverse was true for the Specific Risk/Need score (AUC = .70 any reoffending; AUC = .73 violent reoffending).
- Dyck, Campbell and Wershler (2018)
 - 136 community-based offenders (101 male, 35 female), Canada, reoffending, 3.42 years mean follow-up period
 - Total score demonstrated high predictive accuracy for general recidivism for both male (AUC = .75) and female offenders (AUC = .94).
- Jimenez et al. (2018)
 - 19,344 offenders on probation (14,895 male, 4,449 female), US, reoffending, 5.5 year follow-up period
 - No significant interaction between the LS/CMI risk level and gender. LS/CMI predicted risk levels equally well for women and men.
- Olver, Stockdale and Wormith (2014)
 - Meta-analysis of 128 studies
 - Predictive accuracy of total score for male and female offenders was very similar for both general and violent recidivism. For general recidivism, the weighted mean correlation for females was $r = .31$ (random effects, 95% confidence interval = .26 to .35; $k = 45$) and $r = .30$ for males (random effects, 95% confidence interval = .27 to .34; $k = 80$). For violent recidivism, the weighted mean correlation for females was $r = .26$ (random effects, 95% confidence interval = .20 to .32; $k = 12$) and $r = .24$ for males (random effects, 95% confidence interval = .20 to .27; $k = 30$). Substance abuse and personal/emotional subscales had significantly larger effect sizes for females in the prediction of general recidivism.
- Rettinger and Andrews (2010)
 - 411 female prisoners, Canada, reconviction, 57 month follow-up period
 - High predictive accuracy for both general and violent recidivism (AUC = .87 and .86 respectively).

Level of Service/Case Management Inventory (LS/CMI) and Level of Service/Risk Need Responsivity (LS/RNR)

Applicability: Indigenous Offenders & CALD Offenders

Summary



There is no evidence regarding the applicability of the LS/CMI and the LS/RNR to Aboriginal and Torres Strait Islander offenders or CALD offender groups in Australia. Most international research has focused on Canadian Indigenous offenders, and Black and Hispanic/Latino offenders in the US. While these studies generally support the use of the LS/CMI and LS/RNR with various racial and ethnic minority offender groups, some subscales (e.g. substance abuse) appear to be less predictive for Indigenous offenders and CALD offenders.

Australian Research

No empirical evidence available.

International Research

- Gutierrez et al. (2013)
 - Meta-analysis of 32 reports and 12 data sets

For Indigenous offenders, General Risk/Need factors attained small to moderate mean random effect sizes for general recidivism (ranging from $d = .19$ (Family/Marital) (95% confidence interval = .13 to .26, $k = 26$) to $d = .56$ (Criminal History) (95% confidence interval = .46 to .65, $k = 24$). While predictive of both general and violent recidivism, the subscales Criminal History, Substance Abuse and Antisocial Pattern demonstrated significantly lower predictive accuracy for Indigenous than non-Indigenous offenders. For both Indigenous and non-Indigenous offenders, the tool predicted general recidivism better than violent recidivism.
- Jimenez et al. (2018)
 - 19,344 offenders on probation (4,449 racial and ethnic minority), US, reoffending, 5.5 year follow-up period

General Risk/Need score predicted reoffending equally well for White ($r = .21$) and racial and ethnic minority offenders ($r = .21$). Racial and ethnic minority offenders scored higher on all subscales except Substance Abuse.
- Olver, Stockdale and Wormith (2014)
 - Meta-analysis of 128 studies

Weighted mean correlations of $r = .30$ for Indigenous offenders (random effects, 95% confidence interval = .27 to .31, $k = 13$), $r = .32$ for Asian offenders (random effects, 95% confidence interval = .25 to .38, $k = 4$), $r = .30$ for Black offenders (random effects, 95% confidence interval = .16 to .42, $k = 9$), and $r = .21$ for Hispanic/Latino offenders (random effects, 95% confidence interval = .01 to .41, $k = 6$) across a wide range of studies of general recidivism. Within-studies comparisons generated weighted mean correlations of $r = .28$ for White offenders (random effects, 95% confidence interval = .23 to .34, $k = 23$) and $r = .29$ for Indigenous, Asian, Black and Hispanic/Latino offenders (random effects, 95% confidence interval = .23 to .34, $k = 29$), indicating that ethnicity was not a substantive source of effect size variability.
- Wormith, Hogg and Guzzo (2015)
 - 1,692 Indigenous offenders (1,274 male, 418 female), 24,758 non-Indigenous offenders, Canada, reoffending within four years post-release

High predictive accuracy for general recidivism amongst Indigenous offenders (AUC = .72), although slightly higher for non-Indigenous offenders (AUC = .75). Low predictive validity for violent recidivism (AUC = .64) compared to non-Indigenous offenders (AUC = .74). Subscales Education/Employment, Companions and Substance Abuse were significantly less predictive of violent recidivism amongst Indigenous offenders.

Level of Service/Case Management Inventory (LS/CMI) and Level of Service/Risk Need Responsivity (LS/RNR)

Applicability: Mentally Disordered Offenders

Summary



There is no evidence regarding the applicability of the LS/CMI and LS/RNR to mentally disordered offenders in Australia. International research (predominantly in Canada) generally supports the use of the LS/CMI and LS/RNR with mentally disordered offenders, with some indications that the General Risk/Need score is a stronger predictor of general reoffending but less predictive of violent recidivism.

Australian Research

No empirical evidence available.

International Research

- Canales et al. (2014)
 - 138 community-based mentally disordered offenders (99 male, 39 female), Canada, reoffending, mean follow-up period four years
 - General Risk/Need score demonstrated high predictive accuracy for any reoffending (AUC = .77), while violent reoffending was best predicted by Specific Risk/Need score (AUC = .65). General Risk/Need score demonstrated low predictive accuracy for violent reoffending (AUC = .61), while Specific Risk/Need score demonstrated low predictive accuracy for any reoffending (AUC = .63).
- Girard and Wormith (2004)
 - 169 male mentally disordered offenders, Canada, reconviction, 2.5 year follow-up period
 - General Risk/Need score demonstrated high predictive validity for general recidivism (AUC = .73) and moderate predictive validity for violent recidivism (AUC = .68). Specific Risk/Need score demonstrated high predictive accuracy for violent recidivism (AUC = .71) but low predictive accuracy for general recidivism (AUC = .62). Subscale Criminal History was the best predictor of violent recidivism.
- Olver and Kingston (2019)
 - 604 male forensic psychiatric patients, Canada, reincarceration, 1.9 year follow-up period
 - General Risk/Need score was moderately predictive of general recidivism (AUC = .67) but less predictive for violent recidivism (AUC = .58). Subscale Criminal History was the strongest predictors of both general and violent recidivism (AUC = .70 and .62 respectively), with all other subscales demonstrating low predictive accuracy. Specific Risk/Need score demonstrated low predictive accuracy for general recidivism (AUC = .57) but, contrary to previous studies, was not predictive of violent recidivism.

Applicability: Offenders with Intellectual Disability

Summary



There is no evidence, in Australia or internationally, regarding the predictive validity of the LS/CMI and LS/RNR amongst offenders with a cognitive impairment or intellectual disability.

Australian Research

No empirical evidence available.

International Research

No empirical evidence available.

Level of Service/Case Management Inventory (LS/CMI) and Level of Service/Risk Need Responsivity (LS/RNR)

Contribution to Risk Practice

- The LS/CMI aids the assessor in identifying risk, need and responsivity factors relevant to the individual's likelihood of reoffending and of other issues relevant to a holistic case management plan.
- Many of the factors identified within the assessment can act as targets for treatment/change and the tool can aid assessors in determining the level of monitoring and supervision required with regard to the formulation of case management plans.
- The LS/CMI has an ability to highlight the strengths of the individual. These are factors that would actively enable the individual to desist from further offending and enables assessors to override the degree of service provision generated from Section 1 on the basis of individual strengths.
- Assessors are given the chance to elaborate on factors which have been highlighted as a strength in the 'General Risk/Needs' section (Andrews, Bonta & Wormith, 2004).
- Research literature also describes how the LS/CMI may be used in recommendations for sentencing (see Wolbransky, Serico & Heilbrun, 2012).

Other Considerations

- The LS/CMI was pilot tested as the LSI-OR (Andrews, Bonta & Wormith, 1995) for a number of years prior to its publication in 2004.
- Translations into a country's native language have demonstrated the reliability and predictive validity of the LS/CMI in Punjab, Pakistan (Bhutta & Wormith, 2016) and amongst French speakers in Canada (Guay, 2016).
- While the LS/CMI allows for a professional override of degree of service provision based on an assessment of offender strengths and specific risk factors that are not captured in Section 1, research shows that use of a professional override reduces the predictive validity of LS assessments (Guay & Parent, 2018; Wormith, Hogg & Guzzo, 2012). Accordingly, assessors should use overrides cautiously and sparingly, and only in regard to conditions and circumstances that are well defined and have empirical support.
- Independent research (not author-affiliated) has examined the predictive validity of the LS scales among special offender groups, including gang and non-gang members (Guay, 2012); driving while impaired offenders (Pilon, Jewell & Wormith, 2015) and child exploitation material offenders (Pilon, 2016). Generally, these studies have found that LS assessments do not perform as well for predicting more specific recidivism outcomes as they do with more general measures of outcome, nor as well as specialty assessments that were designed to predict specific types of antisocial behaviour. Studies have also examined the effect of treatment on change in LS/CMI scores (e.g. Holliday, Heilbrun & Fretz, 2012), finding that the LS/CMI is sensitive to change over time and has utility as a means of measuring improvement in dynamic needs areas.
- The assessor should be aware that the LS/CMI is not a comprehensive measure of mitigating and aggravating factors that contribute to offender risk practices (Andrews, Bonta & Wormith, 2004).

Violence Risk Assessment

Historical Clinical Risk-20 Version 2 (HCR-20^{V2})

Webster, Douglas, Eaves and Hart (1997)

Description

- The HCR-20 is a 20-item clinical guide for the structured assessment of violence risk intended for use with civil psychiatric, community, forensic, and criminal justice populations. It was originally published in 1995 but was updated in 1997 (HCR-20^{V2}) and, more recently, in 2013 (HCR-20^{V3}). It is one of the most commonly used structured clinical judgment tools (Singh et al., 2014).
- The instrument is comprised of ten historical variables ('H' Scale) (e.g. previous violence, relationship difficulties, past problems with substance use or employment, trauma history), including a personality disorder item. There are also five clinical variables ('C' Scale) (e.g. insight, violent ideation, symptoms of major mental illness) and five items describing areas relevant to future risk management ('R' Scale) (e.g. future plans for housing, presence of social supports, treatment/supervision response) encompassing relevant past, present, and future considerations.
- Each item may be scored on a three-point scale ('not present'; 'possibly or partially present'; 'present'). A final summary risk rating (SRR) of low, moderate or high is then formulated based on a clinical evaluation of all relevant information.
- Given the differences between the two most recent versions of the HCR-20, the empirical evidence underpinning the utility of the HCR-20^{V2} and HCR-20^{V3} will be presented separately.

Strengths

- The HCR-20^{V2} has the capacity to guide clinical judgment about intervention and risk management (Gray, Taylor & Snowden, 2008).
- Large research base that supports the utility of the HCR-20^{V2} total scores in predicting violence and recidivism across multiple samples and types of settings.

Empirical Grounding

Summary



The HCR-20^{V2} is founded on strong empirical research.

- Research has shown that the HCR-20^{V2} includes static and dynamic factors that have sound empirical grounding (Douglas et al., 2011).

Inter-Rater Reliability

Summary



Research, both in Australia and internationally, has consistently found that the HCR-20^{V2} demonstrates a high level of inter-rater reliability.

Australian Research

- Shepherd, Campbell and Ogloff (2017)
 - Forensic psychiatric patients, VictoriaHigh inter-rater reliability for total scores (ICC = .93) and subscales (Historical ICC = .95; Clinical ICC = .86; Risk Management ICC = .86).

International Research

- Douglas, Ogloff and Hart (2003)
 - Forensic psychiatric patients, CanadaHigh inter-rater reliability for total scores (ICC = .85) and Historical and Clinical subscales (ICC = .90 and .79 respectively). Risk Management subscale achieved only moderate inter-rater reliability (ICC = .49). Inter-rater reliability was greater for total scores than SRRs (ICC = .61).
- Gray, Taylor and Snowden (2008)
 - Male forensic psychiatric patients, EnglandHigh inter-rater reliability for total scores (ICC = .80) and subscales (Historical ICC = .92; Clinical ICC = .90; Risk Management ICC = .85).
- Mills, Kroner and Hemmati (2007)
 - Male prisoners, CanadaHigh inter-rater reliability for total scores (ICC = .85).

General Predictive Validity

Summary



While there is no research regarding the predictive validity of the HCR-20^{V2} in Australian non-mentally disordered offender populations, a large body of international research across various jurisdictions provides strong evidence of its utility in predicting violent reoffending amongst general offenders, with both total scores and subscale scores demonstrating moderate to high predictive validity.

Australian Research

No empirical evidence available.

International Research

- Coid et al. (2009)
 - 1,353 male prisoners convicted of a sexual or violent offence, England, reconviction in two years post-release
 - Total score demonstrated moderate predictive accuracy for any reconviction (AUC = .67), violent reconvictions (AUC = .67) and acquisitive reconvictions (AUC = .69).
- Dahle (2006)
 - 307 male prisoners, Germany, violent reoffending in 10 years post-release
 - Total score demonstrated high predictive accuracy (AUC = .71).
- Douglas, Yeomans and Boer (2005)
 - 188 male prisoners convicted of a violent offence, Canada, violent recidivism post-release, mean follow-up period 7.7 years
 - Total score, as well as all three subscales, demonstrated high predictive accuracy (total score AUC = .82; Historical AUC = .72; Clinical AUC = .79; Risk Management AUC = .80). Summary risk rating also demonstrated high predictive accuracy (AUC = .78).
- Mills, Kroner and Hemmati (2007)
 - 83 male prisoners convicted of a violent offence, Canada, violent reoffending post-release, mean follow-up period 4.6 years
 - Total score demonstrated high predictive accuracy (AUC = .72), as did the Clinical and Risk Management subscales (AUC = .75 and .71 respectively). Historical subscale demonstrated moderate predictive accuracy (AUC = .67). Clinical items were most strongly related to violence.
- Neves, Gonçalves and Palma-Oliveira (2011)
 - 158 offenders on probation or parole (137 male, 21 female), Portugal, reoffending, mean follow-up period 13 months
 - Total score demonstrated high predictive accuracy for both general and violent reoffending (AUC = .84 and .81 respectively). as did the Historical and Risk Management subscales (Historical AUC = .79 (general) and .83 (violent); Risk Management = .80 (general) and .72 (violent)). Clinical subscale demonstrated high predictive accuracy for general reoffending (AUC = .79) but moderate predictive accuracy for future violence (AUC = .69).

Applicability: Female Offenders

Summary



International research indicates that the HCR-20^{V2} has some predictive value amongst female offenders, demonstrating low to moderate predictive accuracy for reconviction, and high predictive accuracy for institutional violence. Of the subscales, the Clinical subscale demonstrates the strongest predictive accuracy for recidivism. There are no studies examining the predictive validity of the HCR-20^{V2} in Australian female offender populations.

Australian Research

No empirical evidence available.

International Research

- Coid et al. (2009)
 - 304 female prisoners convicted of sexual or violent offence, England, reconviction in two years post-release
 - Total score demonstrated moderate predictive accuracy for any reconviction (AUC = .67) and violent reconvictions (AUC = .70). Low predictive accuracy for acquisitive reconviction (AUC = .61).
- de Vogel, Bruggeman and Lancel (2019)
 - 78 female forensic psychiatric inpatients, Netherlands, reconviction, mean follow-up period 11.8 years
 - Total score demonstrated low predictive accuracy for both general (AUC = .63) and violent reconviction (AUC = .59), as did the Risk Management subscale (AUCs = .59 and .59). Both types of reconviction were best predicted by the Clinical subscale (AUC general = .66; AUC violent = .64) which also demonstrated the strongest predictive accuracy for general reconviction (AUC = .68) in a shorter three year follow-up period. Historical subscale achieved AUC values marginally above chance (AUC general = .51; AUC violent = .52). HCR-20^{V3} was found to yield more significant AUC values when compared to the HCR-20^{V2}.
- Eisenbarth et al. (2012)
 - 80 female offenders, Germany, reconviction, mean follow-up period 8 years
 - Did not demonstrate predictive accuracy for reconviction.
- Warren et al. (2017)
 - 183 female offenders, US, institutional violence, unknown follow-up period.
 - Total score demonstrated high predictive accuracy for retrospective reports of institutional violence (AUC = .74), as did the Historical subscale (AUC = .74). Clinical and Risk Management subscales demonstrated moderate predictive accuracy for retrospective reports of institutional violence (AUC = .67 and .63 respectively).

Applicability: Indigenous Offenders & CALD Offenders

Summary



There is no evidence regarding the applicability of HCR-20^{v2} to Aboriginal or Torres Strait Islander offenders or CALD offender populations in Australia. International studies of various Indigenous offender and CALD offender groups suggests that the HCR-20^{v2} performs similarly across cultural groups.

Australian Research

No empirical evidence available.

International Research

- Arai et al. (2016)
 - 127 forensic psychiatric inpatients (108 male, 19 female), Japan, institutional violence, 3 and 6 months follow-up period
 - Total score and Clinical and Risk Management subscales, demonstrated high predictive accuracy at both 3 months (AUC = .84, .90 and .85 respectively) and 6 months (AUC = .80, .77 and .79 respectively). The Historical subscale had weak predictive ability (AUC 3 months = .58; AUC 6 months = .62). Authors concluded that the predictive accuracy of the HCR-20 in Asian samples may be similar to Western countries despite ethnic and cultural differences.
- Fujii et al. (2005)
 - 108 forensic psychiatric patients (88 male, 20 female), US, institutional violence, six month follow-up period
 - Total score achieved moderate to high AUC values amongst Hawaiian Native and European American groups (.73 and .64 respectively). Predictive accuracy for Asian American individuals was low (AUC = .58). No significant differences in AUC values across racial and ethnic groups.
- Snowden, Gray and Taylor (2010)
 - 1,182 mentally disordered offenders discharged from medium secure psychiatric facilities (74% White, 22% African Caribbean, 4% 'other'), England, reconviction in two years post-discharge
 - Moderate to high AUCs for White and African Caribbean offenders (.72 and .66 respectively). Historical and Risk Management subscales showed similar results (AUC = .70 and .69 respectively) but Clinical subscale achieved AUC value marginally above chance (AUC = .54).

Applicability: Mentally Disordered Offenders

Summary



There is some evidence supporting the utility of the HCR-20^{v2} in predicting reconviction amongst Australian mentally disordered offenders, with one study finding that the total score demonstrates high predictive accuracy. International evidence for the predictive validity of the HCR-20^{v2} in this offender population is mixed, ranging from low to high across various jurisdictions.

Australian Research

- Shepherd, Campbell and Ogloff (2017)
 - 136 forensic psychiatric patients (98 male, 38 female), Victoria, reconviction, four year follow-up period
 - Total score demonstrated high predictive accuracy for any reconviction (AUC = .76) and moderate predictive accuracy for a violent reconviction (AUC = .68). Predictive accuracy of Historical and Risk Management subscales was high for any reconviction and moderate for violent reconvictions. Clinical subscale did not significantly predict any form of reconviction and achieved AUC values marginally above chance.

International Research

- Dolan and Fullam (2007)
 - 136 male forensic psychiatric inpatients England, institutional violence, 12 months follow-up period
 - Total score demonstrated high predictive accuracy (AUC = .71).
- Douglas, Ogloff and Hart (2003)
 - 100 insanity acquittees (91 male, 9 female), Canada, violent act post-discharge (reconviction, self-report or collateral report), mean follow-up period 3.5 years
 - Total score demonstrated moderate predictive accuracy (AUC = .67), as did the Clinical subscale (AUC = .68). Historical and Risk Management subscales demonstrated low predictive accuracy (AUC = .63 and .53 respectively). SRRs demonstrated the highest observed predictive accuracy (AUC = .69).
- Jeandarme et al. (2017)
 - 105 insanity acquittees, Belgium, violent behaviour on conditional release, mean follow-up period 26 months
 - Total score demonstrated low predictive accuracy (AUC = .60), as did all three subscales (Historical AUC = .57; Clinical AUC = .57; Risk Management AUC = .63). Non-recidivists (low risk) individuals were identified with higher accuracy than those assessed as high risk.
- McDermott, Dualan and Scott (2011)
 - 146 forensic psychiatric inpatients (126 male, 20 female), US, institutional violence, 20 week follow-up period
 - Total score demonstrated high predictive accuracy (AUC = .73).
- Ramesh et al. (2018)
 - Meta-analysis of 27 studies, institutional violence
 - Total score demonstrated moderate predictive accuracy (AUC = .70).
- Snowden, Gray and Taylor (2010)
 - 1,182 mentally disordered offenders discharged from medium secure psychiatric facilities (gender breakdown not reported), England, reconviction in two years post-discharge
 - Total score demonstrated high predictive accuracy (AUC = .71).

Applicability: Offenders with Intellectual Disability

Summary



There is good research evidence, both in Australia and internationally, that the HCR-20^{V2} performs well when predicting recidivism amongst offenders with a cognitive impairment or intellectual disability, with studies consistently finding that the total score demonstrates high predictive accuracy for both reoffending and institutional violence in this offender population.

Australian Research

- Verbrugge, Goodman-Delahunty and Frize (2011)
 - 59 community-based offenders with ID (55 male, 4 female), NSW, reoffending, two year follow-up periodTotal score and SRRs demonstrated high predictive accuracy for both violent (AUC = .80 and .81 respectively) and general (AUC = .94 and .88 respectively) recidivism. AUCs for general recidivism were larger than those for violent recidivism, for all three subscales.

International Research

- Fitzgerald et al. (2013)
 - 25 offenders with ID (23 male, 2 female), England, institutional aggression, six month follow-up periodTotal score demonstrated high predictive accuracy for any physical aggression (AUC = .77) and severe physical aggression (AUC = .79).
- Gray, Taylor and Snowden (2011)
 - 115 male forensic psychiatric patients with ID, England, reconviction in 2 years post-dischargeTotal score demonstrated high predictive accuracy for violent reconviction and any reconviction (AUCs = .80). Historical scale demonstrated highest predictive accuracy for violent reconvictions (AUC = .84), followed by Risk subscale (AUC = .70) and Clinical subscale (AUC = .68).
- Lindsay et al. (2008)
 - 212 male forensic psychiatric patients with ID across high, medium and low secure settings, England and Wales, institutional violence, follow-up period not reportedTotal score demonstrated high predictive accuracy (AUC = .72). Subscales all demonstrated predictive accuracy in the low to moderate range (Historical AUC = .68; Clinical AUC = .67; Risk Management AUC = .62).
- Morrissey et al. (2007)
 - 60 male offenders with ID, England and Wales, aggression, 12 month follow-up periodTotal score demonstrated moderate predictive accuracy for interpersonal physical aggression (AUC = .68) and high predictive accuracy for verbal/property aggression (AUC = .77).

Historical Clinical Risk-20 Version 2 (HCR-20^{V2})

Contribution to Risk Practice

- The HCR-20 can identify a number of risk and responsivity factors relevant to the individual's risk of violent recidivism.
- Many of the factors identified by the tool can act as targets for treatment/change (e.g. insight, relationship factors) and the instrument can aid decisions regarding the level of monitoring and supervisory strategies, in relation to individuals who pose minimal to high levels of risk for recidivism.
- The HCR-20 can aid assessors in developing risk formulations and risk management strategies.

Other Considerations

- The HCR-20 should be completed using information obtained from interviews with the individual and collateral information.
- The HCR-20^{V2} draws on Psychopathy Checklist-Revised (PCL-R) rating scores as part of the predictive measurement. Thus, each HCR-20^{V2} assessment must be accompanied by an assessment of the PCL-R or an existing file score.
- The time period for which an assessment is produced needs to be considered. For example, Snowden and colleagues (2007) found that the Clinical subscale is a good predictor of institutional violence over a three month follow-up period but a poor predictor of reconviction over a period of several years.
- The dynamic items (i.e. Clinical and Risk Management subscales) are capable of indexing change, as they are dependent on current functioning and context, and can act as a risk barometer. In addition, some of the Historical items may not necessarily be 'fixed' (e.g. changes in the offender's relationship or employment status) (Douglas et al., 2001).
- Few studies have examined the predictive validity of the categorical risk ratings generated by the HCR-20^{V2} (low, moderate, high), with research tending to focus on the validity of the numerical risk scores generated by the scales. This is despite test developers recommending that the categorical risk ratings be used in clinical practice (Webster et al., 1997). Research that has examined the validity of the categorical risk ratings suggests that they demonstrate moderate-to-high predictive validity, with some studies finding that the risk ratings add incremental predictive validity to the numerical scores (de Vogel et al., 2004; Douglas, Ogloff, & Hart, 2003; Douglas, Yeomans, & Boer, 2005).

Historical Clinical Risk-20 Version 3 (HCR-20^{V3})

Douglas, Hart, Webster and Belfrage (2013)

Description

- The HCR-20^{V3} represents the latest iteration of the HCR-20 and builds upon the foundation laid by previous versions.
- The HCR-20^{V3} retains many of the core features of the HCR-20^{V2} but changes have been made to the structure to improve the application of the instrument in clinical practice and to modify and revise items to reflect an updated analysis of the research literature.
- The primary changes include changes to the names and content of some basic risk factors. For example, the 'Personality Disorder' risk item in the H scale has been broadened to include personality disorders other than psychopathy and the PCL-R assessment is no longer required to score the presence of psychopathy. Other changes to the risk factors include the addition of sub-items for complex risk factors and 'indicators' for each item and sub-item which provide examples of specific ways in which a given risk factor might manifest at the individual level. The Numerical item ratings (0, 1, 2) have also been replaced with nominal ratings (N = Not Present, P = Possibly or Partially Present, or Y = Present).
- The addition of 'Relevance ratings' is another key change introduced by the HCR-20^{V3}. These ratings allow the assessor to consider, not only the presence of risk factors, but also their causal importance to a person's risk for violence. The inclusion of 'relevance ratings' (also coded on a three-point scale) emphasises that risk factors are not equally relevant to all persons who possess them.
- The HCR-20^{V3} also adds two new summary risk ratings for Serious Physical Harm and Imminence of Harm (risk of violence over the short-term), in addition to the existing categorical risk rating for overall risk for future violence (Case Prioritisation/Future Violence).

Strengths

- The HCR-20^{V3} encourages the assessment of not only the presence of each risk factor, but also how relevant each one is to an individual's propensity for violence which may assist in the development of future risk management strategies.

Empirical Grounding

Summary



The HCR-20^{V3} is founded on strong empirical research.

- Previous research has shown that the HCR-20 includes static and dynamic factors that have sound empirical grounding (Douglas et al., 2001). The HCR-20^{V3} has been revised to reflect an updated analysis of the research literature since the publication of the HCR-20^{V2}.
- There is an assumption that research regarding the HCR-20^{V2} is applicable to the HCR-20^{V3}, given that they are constitutionally comparable. While research specific to the HCR-20^{V3} is limited, studies have found strong correlations between the two versions (see e.g. de Vogel et al., 2014; Strub, Douglas & Nicholls, 2014; Douglas & Belfrage, 2014; Bjørkly, Eidhammer & Selmer, 2014; Judges, Egan & Broad, 2016).

Historical Clinical Risk-20 Version 3 (HCR-20^{V3})

Inter-Rater Reliability

Summary



While there is no Australian research examining the inter-rater reliability of the HCR-20^{V3}, international studies provide consistent evidence of high inter-rater reliability.

Australian Research

No empirical evidence available.

International Research

- Douglas and Belfrage (2014)
 - Forensic psychiatric patients, Sweden
 - High inter-rater reliability for total scores (ICC = .94), Historical and Clinical subscales (Historical ICC = .94; Clinical ICC = .86) and summary risk rating (ICC = .81). Inter-rater reliability for Risk Management subscale was moderate (ICC = .69).
- Doyle et al. (2014)
 - Forensic psychiatric patients, England
 - High inter-rater reliability for total scores (ICC = .92) and subscales (Historical ICC = .91; Clinical ICC = .90; Risk Management ICC = .93).
- Persson et al. (2017)
 - Forensic psychiatric patients, Sweden
 - High inter-rater reliability for total scores (ICC = .81) and summary risk rating (ICC = .80).
- Smith et al. (2014)
 - Male prisoners, US
 - High inter-rater reliability for Historical subscale (Historical ICC = .92). Inter-rater reliability for Clinical and Risk Management subscales were moderate (Clinical ICC = .67; Risk Management ICC = .68).

General Predictive Validity

Summary



There is limited research examining the predictive validity of the HCR-20^{V3} in general offender populations. However, one small Canadian study found that both the total score and subscale scores demonstrate high predictive accuracy for future violent behaviour.

Australian Research

No empirical evidence available.

International Research

- Strub, Douglas and Nicholls (2014)
 - 56 prisoners (33 male, 23 female) sentenced to less than two years, Canada, violent behaviour 4-6 weeks and 6-8 months post-discharge
 - Total score demonstrated moderate predictive accuracy for violent behaviour 4-6 weeks post-discharge (AUC = .70) and high predictive accuracy for violent behaviour 6-8 months post-discharge (AUC = .79). Summary risk ratings demonstrated high predictive accuracy for violent behaviour 4-6 weeks post-discharge (AUC = .72) and moderate predictive accuracy for violent behaviour 6-8 months post-discharge (AUC = .68). All subscales demonstrated high predictive accuracy for violent behaviour 6-8 months post-discharge (Historical AUC = .74; Clinical AUC = .79; Risk Management AUC = .74).

Historical Clinical Risk-20 Version 3 (HCR-20^{V3})

Applicability: Female Offenders

Summary



International research indicates that the HCR-20^{V3} has some predictive value amongst female offenders, demonstrating low to moderate predictive accuracy for reconviction and institutional violence. Of the subscales, the Clinical subscale demonstrates the strongest predictive accuracy. Note though that studies have involved small and selective samples. There are no studies examining the predictive validity of the HCR-20^{V3} in Australian female offender populations.

Australian Research

No empirical evidence available.

International Research

- de Vogel, Bruggeman and Lancel (2019)
 - 78 female forensic psychiatric inpatients, Netherlands, reconviction, mean follow-up period 11.8 years
 - Total score demonstrated moderate predictive accuracy for both general and violent reconviction (AUCs = .67). Both types of reconviction were best predicted by the Clinical subscale (AUC = .67 (general); AUC = .69 (violent)) which also demonstrated the strongest predictive accuracy in a shorter three year follow-up period. Risk Management subscale demonstrated moderate predictive accuracy for general reconviction (AUC = .64) but low predictive accuracy for violent reconviction (AUC = .61). Predictive accuracy of the Historical Scale was low for general reconviction (AUC = .63) but moderate for violent reconviction (AUC = .64). HCR-20^{V3} was found to yield more significant AUC values when compared to the HCR-20^{V2}.
- Green et al. (2016)
 - 24 female insanity acquittees, US, institutional violence, mean follow-up period 15.5 months
 - Similar correlations between institutional violence and total score ($r = .27$), Historical subscale ($r = .28$) and Clinical subscale ($r = .31$). Although not statistically significant, the Risk Management subscale was negatively associated with institutional violence (ie. there was a trend towards lower scores corresponding with violence) ($r = -.08$).
- Strub, Douglas and Nicholls (2014)
 - 106 offenders and civil psychiatric patients (63 male, 53 female), Canada, violent behaviour 4-6 weeks and 6-8 months post-discharge
 - Total score and summary risk ratings at both 4-6 weeks and 6-8 months post-discharge were not moderated by gender.

Applicability: Indigenous Offenders & CALD Offenders

Summary



There is no research evidence, either in Australia or internationally, regarding the predictive validity of the HCR-20^{V3} amongst Indigenous offenders and CALD offenders.

Australian Research

No empirical evidence available.

International Research

No empirical evidence available.

Historical Clinical Risk-20 Version 3 (HCR-20^{V3})

Applicability: Mentally Disordered Offenders

Summary



International evidence for the applicability of the HCR-20^{V3} to mentally disordered offenders is mixed, with findings ranging from low to high predictive validity across different jurisdictions, and across the different subscales and summary risk ratings. The total score tends to demonstrate moderate to high predictive accuracy, while the Historical subscale score tends to be less predictive of future violence when compared to the other subscales. There is no research examining the predictive validity of the HCR-20^{V3} in Australian mentally disordered offender populations.

Australian Research

No empirical evidence available.

International Research

- Coid et al. (2015)
 - 788 forensic psychiatric patients (344 male, 43 female), England, violent act six months post-discharge
 - Clinical and Risk Management subscales demonstrated moderate predictive accuracy (AUCs = .67). Historical subscale demonstrated low predictive accuracy (AUC = .60).
- de Vogel, van den Broek, and de Vries Robbé (2014)
 - 86 male forensic psychiatric patients, Netherlands, violent reoffending, 12, 24 and 36 months post-discharge
 - Total score demonstrated high predictive accuracy at both 12 and 24 months post-discharge (AUC = .77 and .75 respectively).
 - Predictive accuracy for violent recidivism three years post-discharge was moderate (AUC = .67). Summary risk ratings demonstrated high predictive accuracy at 12 months post-discharge (AUC = .72) and moderate predictive accuracy at 24 and 36 months post-discharge (AUC = .67 and .64 respectively).
- Doyle et al. (2014)
 - 387 forensic psychiatric patients (365 male, 22 female), England and Wales, violent act six and 12 months post-discharge
 - Moderate to high predictive accuracy at both six and 12 months post-discharge for total score (AUC = .73 and .70) and Clinical subscale (AUC = .75 and .71). Historical subscale demonstrated low predictive accuracy at six and 12 months post-discharge (AUC = .63 and .63 respectively) while Risk Management subscale demonstrated low to moderate predictive accuracy at six and 12 months post-discharge (AUC = .67 and .63).
- Hogan and Olver (2016)
 - 99 forensic psychiatric inpatients (85 male, 14 female), Canada, institutional violence, mean follow-up period 19 months.
 - High predictive accuracy for total score (AUC = .76) and Clinical and Risk Management subscales (AUCs = .76). Historical subscale demonstrated moderate predictive accuracy (AUC = .64). Summary risk ratings ranged from low predictive accuracy for Serious Harm (AUC = .44) to moderate for Case Prioritisation (AUC = .68) and high for Imminent Violence (AUC = .75).
- Persson et al. (2017)
 - 200 forensic psychiatric patients (174 male, 26 female), Sweden, violent act, 12 month follow-up period
 - High predictive accuracy for total score (AUC = .78) and summary risk rating (AUC = .75).

Applicability: Offenders with Intellectual Disability

Summary



There is no research evidence, either in Australia or internationally, regarding the validity of the HCR-20^{V3} amongst offenders with a cognitive impairment or intellectual disability.

Australian Research

No empirical evidence available.

International Research

No empirical evidence available.

Contribution to Risk Practice

- The HCR-20^{V3} encourages the assessment of not only the presence of each risk factor, but also how relevant each one is to an individual's propensity for violence. In this way, the 'relevance rating' provides guidance regarding the extent to which a given risk factor should be emphasised in violence risk management plans and interventions, and provides an added structure that allows for more transparency in clinical and legal assessments.
- The addition of two new summary risk ratings for Serious Physical Harm and Imminence of Harm allows for a more detailed formulation of violence risk, but research has not yet examined the relationship between these risk ratings and outcomes. For example, a person may be judged to pose a high risk for violence generally, but a low or moderate risk for serious physical harm or for imminent violence.

Other Considerations

- An additional supplement to the HCR-20^{V3} for the assessment of violence in women has been developed (de Vogel et al., 2012). This tool, known as the Female Additional Manual, is comprised of additional guidelines to five of the historical HCR-20 items and nine additional risk items reflecting gender-responsive issues specific to the evaluation of female offenders. While the tool has yet to be widely researched, preliminary findings indicate that the FAM demonstrates good predictive validity for institutional violence and self-destructive behaviour (de Vogel & de Vries Robbé, 2013).
- The authors recommend that the HCR-20^{V3} be repeated (every 6 to 12 months) to take into account changes in circumstances (Douglas et al., 2013).
- Given the strong correlations between version 2 and version 3 of the HCR-20 (see e.g. de Vogel et al., 2014; Strub, Douglas & Nicholls, 2014; Douglas & Belfrage, 2014; Bjørkly, Eidhammer & Selmer, 2014; Judges, Egan & Broad, 2016), research regarding the predictive validity of the HCR-20^{V2} is highly likely to extrapolate to the HCR-20^{V3}. Accordingly, while research specific to the HCR-20^{V3} is limited, practitioners should use the more recent iteration of the tool if they have access to both.

Violence Risk Scale (VRS)

Wong and Gordon (1999)

Description

- The VRS is a 26-item actuarial risk assessment tool designed to assess the risk of violent reoffending for incarcerated individuals and forensic psychiatric patients being considered for community access.
- It can be used to monitor changes in risk and motivation to change.

Strengths

- A discretionary override is available for situations that are not captured by the risk factors found in the tool.
- The tool has an in-built methodology for appraising change based on the Stages of Change model.

Empirical Grounding

Summary



The VRS is founded on strong empirical research.

- The VRS static and dynamic risk factors are deemed to be empirically or theoretically related to violent recidivism (Wong & Gordon, 1999).

Inter-Rater Reliability

Summary



While there is no Australian research examining the inter-rater reliability of the VRS, international research provides consistent evidence of high inter-rater reliability across a range of jurisdictions.

Australian Research

No empirical evidence available.

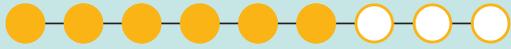
International Research

- Dolan et al. (2008)
 - Forensic male and female psychiatric inpatients, England
 - High inter-rater reliability for composite score, static subscale and dynamic subscale (ICCs= .89, .96 and .85 respectively).
- Lewis, Olver and Wong (2012)
 - Male offenders with psychopathic traits, Canada
 - High inter-rater reliability for total score (ICC values ranging from .82 to .84).
- Wong and Parhar (2011)
 - Male community-based offenders, Canada
 - High inter-reliability for total score (ICC = .93).
- Zhang et al. (2012)
 - Forensic male and female psychiatric inpatients, China
 - High inter-rater reliability for total score (ICC = .80).

Violence Risk Scale (VRS)

General Predictive Validity

Summary



While there is no Australian research regarding the predictive validity of the VRS amongst general offenders, international studies have found that the VRS demonstrates moderate to high predictive accuracy for general and violent reconviction in the general offending population.

Australian Research

No empirical evidence available.

International Research

- Lewis, Olver and Wong (2012)
 - 150 male violent prisoners with psychopathic traits who attended a violence reduction treatment program, Canada, violent reconviction in three years post-release
 - Moderate predictive accuracy for violent reconviction (AUC = .65). Pre-treatment total score was not significant.
- Wong and Parhar (2011)
 - 59 male community-based offenders, Canada, reconviction, seven year follow-up period
 - High predictive accuracy for both violent (AUC = .83) and any reconviction (AUC = .72).
- Wong and Gordon (2006)
 - 918 male prisoners, Canada, reconviction, in 4.4 years post-release
 - High predictive accuracy for any reconviction (AUC = .74), violent reconviction (AUC = .75) and non-violent reconviction (AUC = .72).

Applicability: Female Offenders

Summary



There is no research evidence, either in Australia or internationally, regarding the validity of the VRS amongst female offenders.

Australian Research

No empirical evidence available.

International Research

No empirical evidence available.



Violence Risk Scale (VRS)

Applicability: Indigenous Offenders & CALD Offenders

Summary



There is no research evidence, either in Australia or internationally, regarding the predictive validity of the VRS amongst Indigenous offenders and CALD offenders.

Australian Research

No empirical evidence available.

International Research

No empirical evidence available.

Applicability: Mentally Disordered Offenders

Summary



International research of the predictive validity of the VRS amongst mentally disordered offenders is limited to studies of its ability to predict institutional violence. Results of these studies provide tentative support for the use of the VRS, although a recent meta-analysis found that the VRS demonstrated low predictive accuracy in this offender group. There are no Australian studies examining the predictive validity of the VRS in Australian mentally disordered offender populations.

Australian Research

No empirical evidence available.

International Research

- Dolan and Fullam (2007)
 - 136 male forensic psychiatric inpatients, England, institutional violence in 12 months post-assessment
 - Total score demonstrated high predictive accuracy (AUC = .71) and ability to discriminate between violent and non-violent patients ($d = .72$). Patients who had engaged in institutional violence had higher mean VRS composite and subscale scores than non-violent group.
- Dolan et al. (2008)
 - 147 forensic psychiatric inpatients (136 male, 11 female), England, institutional violence, mean follow-up period 371 days
 - Total score demonstrated moderate predictive accuracy (AUC = .69). Dynamic subscale was a better predictor of institutional violence than static subscale (AUC = .70 and .60 respectively).
- Ramesh, Igoumenou, Vazquez Montes and Fazel (2018)
 - Meta-analysis of four studies, institutional violence
 - Total score demonstrated low predictive accuracy (AUC = .63).

Applicability: Offenders with Intellectual Disability

Summary



There is no research evidence, either in Australia or internationally, regarding the predictive validity of the VRS amongst offenders with a cognitive impairment or intellectual disability.

Australian Research

No empirical evidence available.

International Research

No empirical evidence available.

Violence Risk Scale (VRS)

Contribution to Risk Practice

- The VRS has the ability to create awareness of static risk factors and can prompt further assessment of the risk of reoffending.
- The VRS can aid assessors in identifying risk and responsivity factors (such as treatment responsivity) which can contribute to measuring progress/deterioration in factors linked to the individual's offending behaviours.
- Using the combination of risk and treatment readiness (stage of change) information, a VRS assessment can also inform the levels of monitoring and rehabilitation efforts and other risk management strategies (Wong & Gordon, 2006).
- As the VRS consists of 20 dynamic factors that can be used to assess risk and changes in risk posed by the individual, it can inform treatment targets and management plans, and the re-assessment of risk. The tool can also assist in release decision-making (Daffern, 2007).

Other Considerations

- Few validation studies conducted with female offenders, Indigenous offenders and CALD offender groups.
- The second edition (VRS) was an experimental version so named when it was under development. The content of the VRS and VRS 2nd Edition are essentially the same with minor changes to the wording. The authors suggest that tool users should continue to refer to the instrument as the VRS without reference to specific editions (S Wong, personal communication with the Scottish Risk Management Authority, January 2013).

References

- Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290.
- Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.
- Fleiss, J.L., Levin, B.A., & Paik, M.C. (2003). *Statistical methods for rates and proportions* (3rd ed.) Hoboken, NJ: Wiley-Interscience.
- Helmis, L.M., & Babchishin, K.M. (2017). Primer on risk assessment and the statistics used to evaluate its accuracy. *Criminal Justice and Behavior*, 44(1), 8-25.
- Mossman, D. (2013). Evaluating risk assessments using receiver operating characteristic analysis: Rationale, advantages, insights, and limitations. *Behavioral Sciences and the Law*, 31, 23-39.
- Rice, M.E., & Harris, G.T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, 63, 737-748.
- Rice, M.E., & Harris, G.T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r*. *Law and Human Behavior*, 5, 615-620.
- Risk Management Authority (2019). *Risk Assessment Tools Evaluation Directory* (4th ed). Retrieved from: <http://rma.scot/research/rated/>
- Level of Service Inventory-Revised (LSI-R)**
- Andrews, D.A., & Bonta, J. (1995). *The Level of Service Inventory-Revised*. Toronto, Ontario, Canada: Multi-Health Systems.
- Andrews, D.A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). New Providence, NJ: Matthew Bender.
- Bhutta, M.H., & Wormith, J.S. (2016). An examination of a risk/needs assessment instrument and its relation to religiosity and recidivism among probationers in a Muslim culture. *Criminal Justice and Behavior*, 43, 204-229.
- Campbell, M.A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior*, 36, 567–590.
- Chenane, J.L., Brennan, P.K., Steiner, B., & Ellison, J.M. (2015). Racial and ethnic differences in the predictive validity of the Level of Service Inventory-Revised among prison inmates. *Criminal Justice and Behavior*, 42, 286–303.
- Dahle, K. (2006). Strengths and limitations of actuarial prediction of criminal reoffence in a German prison sample: A comparative study of LSI-R, HCR-20 and PCL-R. *International Journal of Law and Psychiatry*, 29, 431-442.
- Duwe, G., & Rocque, M. (2016). A jack of all trades but a master of none? Evaluating the performance of the Level of Service Inventory-Revised (LSI-R) in the assessment of risk and need. *Corrections*, 1(2), 81-106.
- Fass, T.L., Heilbrun, K., DeMatteo, D., & Frenz, R. (2008). The LSI-R and the COMPAS: Validation data on two risk-needs tools. *Criminal Justice and Behavior*, 35, 1095-1108.
- Georgiou, G. (2019). Weights matter: Improving the predictive validity of risk assessments for criminal offenders. *Journal of Offender Rehabilitation*, 58(2), 92-116.
- Hausam, J., Lehmann, R.J.B., & Dahle, K. (2018). Predicting offenders' institutional misconduct and recidivism: The utility of behavioural ratings by prison officers. *Frontiers in Psychiatry*, 9, 679.
- Hendricks, B., Werner, T., Shipway, L., & Turinetti, G.J. (2006). Recidivism among spousal abusers. *Journal of Interpersonal Violence*, 21, 703-716.
- Hollin, C.R., Palmer, E.J., & Clark, D. (2003). Level of Service Inventory-Revised profile of English prisoners: A needs analysis. *Criminal Justice and Behavior*, 30, 422–440.
- Hsu, C., Caputi, P., & Byrne, M.K. (2009). The Level of Service Inventory-Revised (LSI-R): A useful risk assessment measure for Australian offenders? *Criminal Justice and Behavior*, 36, 728-740.
- Hsu, C., Caputi, P., & Byrne, M.K. (2010). Level of Service Inventory-Revised: Assessing the risk and need characteristics of Australian Indigenous offenders. *Psychiatry, Psychology and Law*, 17, 355-367.
- Labrecque, R., Smith, P., Lovins, B., & Latessa, E. (2014). The importance of reassessment: How change in the LSI-R risk score can improve the prediction of recidivism. *Journal of Offender Rehabilitation*, 53, 116-126.
- Lowenkamp, C.T., Holsinger, A.M., Brusman-Lovins, L., & Latessa, E.J. (2004). Assessing the inter-rater agreement of the Level of Service Inventory-Revised. *Federal Probation*, 68, 34-38.
- Lowenkamp, C.T., Lovins, B., & Latessa, E.J. (2009). Validating the Level of Service Inventory-Revised and the Level of Service Inventory: Screening Version with a sample of probationers. *The Prison Journal*, 89, 192-204.
- Manchak, S.M., Skeem, J.L., & Douglas, K. S. (2008). Utility of the Revised Level of Service Inventory (LSI-R) in predicting recidivism after long-term incarceration. *Law and Human Behavior*, 32, 477-488.
- Manchak, S.M., Skeem, J.L., Douglas, K. S. & Siranosian, M. (2009). Does gender moderate the predictive utility of the Level of Service Inventory-Revised (LSI-R) for serious violent offenders? *Criminal Justice and Behavior*, 36, 425-442.
- Ostermann, M., & Herrschaft, B.A. (2013). Validating the Level of Service Inventory-Revised: A gendered perspective. *The Prison Journal*, 93, 291-312.
- Ostermann, M., & Salerno, L.M. (2016). The validity of the Level of Service Inventory-Revised at the intersection of race and gender. *The Prison Journal*, 96, 554-575.
- Palmer, E J., & Hollin, C.R. (2007). The Level of Service Inventory-Revised with English women prisoners: A needs and reconviction analysis. *Criminal Justice and Behavior*, 34, 971–984.
- Smith, P., Cullen, F.T., & Latessa, E.J. (2009). Can 14,373 women be wrong? A meta-analysis of the LSI-R and recidivism for female offenders. *Criminology & Public Policy*, 8, 183–208.
- Wilson, H.A., & Gutierrez, L. (2014). Does one size fit all? A meta-analysis examining the predictive ability of the Level of Service Inventory (LSI) with Aboriginal offenders. *Criminal Justice and Behavior*, 41, 196–216.

References

- Zhang, J. (2016). Testing the predictive validity of the LSI-R using a sample of young male offenders on probation in Guangzhou, China. *International Journal of Offender Therapy and Comparative Criminology*, 60, 456-468.
- Zhang, J., & Liu, N. (2015). Reliability and validity of the Chinese version of the LSI-R with probationers. *International Journal of Offender Therapy and Comparative Criminology*, 59, 1474-1486.
- Level of Service Inventory-Revised: Screening Version (LSI-R:SV)**
- Andrews, D.A., & Bonta, J. (1998). *The Level of Service Inventory-Revised: Screening Version*. Toronto, ON: Multi-Health Systems.
- Andrews, D.A., Bonta, J., & Wormith, J. (2004). *The Level of Service/Case Management Inventory (LS/CMI): User's Manual*. Toronto, ON: Multi-Health Systems.
- Daffern, M. (2007). The predictive validity and practical utility of structured schemes used to assess risk for aggression in psychiatric inpatient settings. *Aggression and Violent Behavior*, 12, 116-130.
- Daffern, M., Ogloff, J.R., Ferguson, M., & Thomson, L. (2005). Assessing risk for aggression in a forensic psychiatric hospital using the Level of Service Inventory-Revised: Screening Version. *International Journal of Forensic Mental Health*, 4(2), 201-206.
- Ferguson, A.M., Ogloff, J.R.P., & Thomson, L. (2009). Predicting recidivism by mentally disordered offenders using the LSI-R:SV. *Criminal Justice and Behavior*, 36, 5-20.
- Livingston, J.D., Chu, K., Milne, T., & Brink, J. (2015). Probationers mandated to receive forensic mental health services in Canada: Risks/Needs, service delivery, and intermediate outcomes. *Psychology, Public Policy, and Law*, 21(1), 72-84.
- Lowenkamp, C.T., Lovins, B., & Latessa, E.J. (2009). Validating the Level of Service Inventory-Revised and the Level of Service Inventory: Screening Version with a sample of probationers. *The Prison Journal*, 89, 192-204.
- McCafferty, J.T., & Scherer, H. (2017). Beyond recidivism: Exploring the predictive validity of a correctional risk assessment tool on offender victimization. *The Prison Journal*, 97(6), 674-691.
- Walters, G.D. (2011). Predicting recidivism with the Psychological Inventory of Criminal Thinking Styles and Level of Service Inventory-Revised: Screening Version. *Law and Human Behavior*, 35, 211-220.
- Walters, G.D., & Schlauch, C. (2008). The Psychological Inventory of Criminal Thinking Styles and Level of Service Inventory-Revised: Screening Version as predictors of official and self-reported disciplinary infractions. *Law and Human Behavior*, 32, 454-462.
- Yessine, A.K., & Bonta, J. (2006). Tracking high-risk, violent offenders: An examination of the National Flagging System. *Canadian Journal of Criminology and Criminal Justice*, 48(4), 573-607.
- Level of Service/Case Management Inventory (LS/CMI) and Level of Service/ Risk Need Responsivity (LS/RNR)**
- Andrews, D.A., & Bonta, J. (1995). *The Level of Service Inventory-Revised*. Toronto, Ontario, Canada: Multi-Health Systems.
- Andrews, D.A., Bonta, J., & Wormith, J. (2004). *The Level of Service/Case Management Inventory (LS/CMI): User's Manual*. Toronto, ON: Multi-Health Systems.
- Andrews D.A., Bonta J., & Wormith J.S. (2008). *The Level of Service/Risk-Need-Responsivity (LS/RNR): User's Manual*. Toronto, ON: Multi-Health Systems.
- Andrews, D.A., Guzzo, L., Raynor, P., Rowe, R.C., Rettinger, L.J., Brews, A., & Wormith, J.S. (2012). Are the major risk/need factors predictive of both female and male reoffending? A test with the eight domains of the Level of Service/Case Management Inventory. *International Journal of Offender Therapy and Comparative Criminology*, 56, 113-133.
- Canales, D.D., Campbell, A., Wei, R., & Totten, A.E. (2014). Prediction of general and violent recidivism among mentally disordered adult offenders: Test of the Level of Service/Risk-Need-Responsivity (LS/RNR) Instrument. *Criminal Justice and Behavior*, 41, 971-991.
- Dyck, H.L., Campbell, M.A., & Wershler, J.L. (2018). Real-world use of the Risk-Need-Responsivity Model and the Level of Service/Case Management Inventory with community-supervised offenders. *Law and Human Behavior*, 42(3), 258-26.
- Girard, L., & Wormith, J.S. (2004). The predictive validity of the Level of Service Inventory-Ontario Revision on general and violent recidivism among various offender groups. *Criminal Justice and Behavior*, 31, 150-181.
- Gordon, H., Kelty, S.F., & Julian, R. (2015). Psychometric evaluation of the Level of Service/Case Management Inventory among Australian offenders completing community-based sentences. *Criminal Justice and Behavior*, 42, 1089-1109.
- Guay, J-P. (2012). *Predicting recidivism with street gang members 2012-02*. Corrections Research User report. Public Safety Canada. Retrieved from: [http:// www.publicsafety.gc.ca/cnt/rsrcs/pblctns/2012-02-prsgm/index-en.aspx](http://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/2012-02-prsgm/index-en.aspx)
- Guay, J-P. (2016). L'évaluation du risqué et des besoins criminogènes à la lumière des données probantes: une étude de validation de la version française de l'Inventaire de niveau de service et de gestion des cas – LS/CMI, [French validation of the Level of Service/Case Management Inventory – LS/CMI]. *European Review of Applied Psychology/Revue Européenne de Psychologie Appliquée*, 66, 199-210.
- Guay, J-P., & Parent, G. (2018). Broken legs, clinical overrides, and recidivism risk: An analysis of decisions to adjust risk levels with the LS/CMI. *Criminal Justice and Behavior*, 45, 82-100.
- Gutierrez, L., Wilson, H.A., Rugge, T., & Bonta, J. (2013). The prediction of recidivism with Aboriginal offenders: A quantitative and theoretically informed review. *Canadian Journal of Criminology and Criminal Justice*, 55, 55-99.

- Holliday, S.B., Heilbrun, K., & Fretz, R. (2012). Examining improvements in criminogenic needs: The risk reduction potential of a structured re-entry program. *Behavioral Sciences and the Law*, 30, 431-447.
- Jimenez, A.C., Delgado, R.H., Vardsveen, T.C., & Weiner, R.L. (2018). Validation and application of the LS/CMI in Nebraska Probation. *Criminal Justice and Behavior*, 45, 863-884.
- Olver, M.E., & Kingston, D.A. (2019). Discrimination and calibration properties of the Level of Service Inventory–Ontario Revision in a correctional mental health sample. *Criminal Justice and Behavior*, 46(1), 5–23.
- Olver, M.E., Stockdale, K.C., & Wormith, J.S. (2014). Thirty years of research on the level of service scales: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment*, 26, 156–176.
- Pilon, A.J.M. (2016). The predictive validity of general and offense-specific risk assessment tools for child pornography offenders' reoffending. Unpublished master's thesis. University of Saskatchewan, Saskatoon, Canada. Retrieved from: <https://ecommons.usask.ca/bitstream/handle/10388/ETD-2016-01-2414/PILON-THESIS.pdf?sequence=3>
- Pilon, A.J., Jewell, K.M., & Wormith, J.S. (2015). *Impaired drivers and their risk of reoffending*. Report submitted to Public Safety Canada. Saskatoon: Centre for Forensic Behavioural Science & Justice Studies, University of Saskatchewan.
- Rettinger, L.J., & Andrews, D.A. (2010). General risk and need, gender specificity, and the recidivism of female offenders. *Criminal Justice and Behavior*, 37, 29-46.
- Rowe, R.C. (1999). *The prediction or recidivism in a parole sample: An examination of two versions of the Level of Service Inventory*. Unpublished report. Carleton University, Ottawa, ON, Canada.
- Wolbransky, M., Serico, J.M., & Heilbrun, K. (2012). Forensic sentencing evaluations under post-Booker Federal Sentencing Guidelines: Incorporating voluntary substance use. *Journal of Forensic Psychology Practice*, 12, 57-67.
- Wormith, J.S., Hogg, S.M., & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. *Criminal Justice and Behavior*, 39, 1511–1538.
- Wormith, J.S., Hogg, S.M., & Guzzo, L. (2015). The predictive validity of the LS/CMI with Aboriginal offenders in Canada. *Criminal Justice and Behavior*, 42, 281–508.
- Historical Clinical Risk-20 Version 2 (HCR-20^{v2})**
- Arai, K., Takano, A., Nagata, T., & Hirabayashi, N. (2017). Predictive accuracy of the Historical-Clinical-Risk Management 20 for violence in forensic psychiatric wards in Japan. *Criminal Behaviour and Mental Health*, 27, 409-420.
- Coid, J., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Roberts, C., Farrington, D.P., & Rogers, R. (2009). Gender differences in structured risk assessment: Comparing the accuracy of five instruments. *Journal of Consulting and Clinical Psychology*, 2, 337-348.
- Dahle, K. (2006). Strengths and limitations of actuarial prediction of criminal reoffence in a German prison sample: A comparative study of LSI-R, HCR-20 and PCL-R. *International Journal of Law and Psychiatry*, 29, 431-442.
- de Vogel, V., Bruggeman, M., & Lancel, M. (2019). Gender-sensitive violence risk assessment: Predictive validity of six tools in female forensic psychiatric patients. *Criminal Justice and Behavior*, 46(4), 528-549.
- de Vogel, V., de Ruiter, C., Hildebrand, M., Bos, B., & van de Ven, P. (2004). Type of discharge and risk of recidivism measured by the HCR-20: A retrospective study in a Dutch sample of treated forensic psychiatric patients. *International Journal of Forensic Mental Health*, 3, 149-165.
- Douglas, K.S., Ogloff, J.R.P., & Hart, S.D. (2003). Evaluation of a model of violence risk assessment among forensic psychiatric patients. *Psychiatric Services*, 54, 1372-1379.
- Douglas, K.S., Webster, C.D., Hart, S.D., Eaves, D., & Ogloff, J.R.P. (2001). *HCR-20 Violence Risk Management Companion Guide*. Florida: Simon Fraser University.
- Douglas, K.S., Yeomans, M., & Boer, D.P. (2005). Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior*, 32, 479-510.
- Doyle, M., & Dolan, M. (2006). Predicting community violence from patients discharged from mental health services. *British Journal of Psychiatry*, 189, 520-526.
- Eisenbarth, H., Osterheider, M., Nedopil, N., & Stadtland, C. (2012). Recidivism in female offenders PCL-R Lifestyle Factor show predictive validity in a German sample. *Behavioral Sciences and the Law*, 30, 575-584.
- Fitzgerald, S., Gray, N.S., Alexander, R.T., Bagshaw, R., Chesterman, P., Huckle, P., Jones, S.K., Taylor, J., Williams, T., & Snowden, R.J. (2013). Predicting institutional violence in offenders with intellectual disabilities: The predictive efficacy of the VRAG and the HCR-20. *Journal of Applied Research in Intellectual Disabilities*, 26, 384-393.
- Fujii, D.E., Tokioka, A.B., Lichten, A.I., & Hishinuma, E. (2005). Ethnic differences in prediction of violence risk with the HCR-20 among psychiatric inpatients. *Psychiatric Services*, 56, 711-716.
- Gray, N.S., Taylor, J., & Snowden, R.J. (2008). Predicting violent reconvictions using the HCR-20. *British Journal of Psychiatry*, 192, 384-387.

References

- Gray, N.S., Taylor, J., & Snowden, R.J. (2011). Predicting violence using structured professional judgement in patients with different mental and behavioural disorders. *Psychiatry Research*, 187, 248-253.
- Lindsay, W.R., Hogue, T.E., Taylor, J.L., Steptoe, L., Mooney, P., O'Brien, G., Johnston, S., & Smith, A.H.A. (2008). Risk assessment in offenders with intellectual disability: A comparison across three levels of security. *International Journal of Offender Therapy and Comparative Criminology*, 52, 90-111.
- Morrissey, C., Hogue, T., Mooney, P., Allen, C., Johnston, S., Hollin, C., Lindsay, W.R., & Taylor, J.T. (2007). Predictive validity of the PCL-R in offenders with intellectual disability in a high secure hospital setting: Institutional aggression. *Journal of Forensic Psychiatry & Psychology*, 18, 1-15.
- McDermott, B.E., Dualan, I.V., & Scott, C.L. (2011). The predictive ability of the Classification of Violence Risk (COVR) in a forensic psychiatric hospital. *Psychiatric Services*, 62, 430-433.
- Mills, J.F., Kroner, D.G., & Hemmati, T. (2007). The validity of violence risk estimates: An issue of item performance. *Psychological Services*, 4, 1-12.
- Neves, A.C., Gonçalves, R.A., & Palma-Oliveira, J.M. (2011). Assessing risk for violent and general recidivism: A study of the HCR-20 and the PCL-R with a non-clinical sample of Portuguese offenders. *International Journal of Forensic Mental Health*, 10(2), 137-149.
- Ramesh, T., Igoumenou, A., Vazquez Montes, M., & Fazel, S. (2018). Use of risk assessment instruments to predict violence in forensic psychiatric hospitals: a systematic review and meta-analysis. *European Psychiatry*, 52, 47-53.
- Shepherd, S., Campbell, R., & Ogloff, J.R.P. (2017). The Utility of the HCR-20 in an Australian sample of forensic psychiatric patients. *Psychiatry, Psychology and Law*, 25, 273-282.
- Singh, J., Fazel, S., Gueorguieva, R., & Buchanan, A. (2014). Rates of violence in patients classified as high risk by structured risk assessment instruments. *British Journal of Psychiatry*, 204, 180-187.
- Snowden, R.J., Gray, N.S., & Taylor, J. (2010). Risk Assessment for Future Violence in Individuals from an Ethnic Minority Group. *International Journal of Forensic Mental Health*, 9, 118-123.
- Snowden, R.J., Gray, N.S., Taylor, J., & MacCulloch, M.J. (2007). Actuarial prediction of violent recidivism in mentally disordered offenders. *Psychological Medicine*, 37, 1539-1550.
- Verbrugge, H.M., Goodman-Delahunty, J., & Frize, M.C.J. (2011). Risk assessment in intellectually disabled offenders: Validation of the suggested ID supplement to the HCR-20. *International Journal of Forensic Mental Health*, 10, 83-91.
- Warren, J.I., Wellbeloved-Stone, J.M., Dietz, P.E., & Millspaugh, S.B. (2017). Gender and Violence Risk Assessment in Prisons. *Psychological Services*. Advance Online Publication.
- Webster, C.D., Douglas, K.S., Eaves, D., & Hart, S.D. (1997). *HCR-20: Assessing risk for violence (version 2)*. Vancouver, Canada: Simon Fraser University.
- Historical Clinical Risk-20 Version 3 (HCR-20^{V3})**
- Bjørkly, S., Eidhammer, G., & Selmer, L. (2014). Concurrent validity and clinical utility of the HCR-20V3 compared with the HCR-20 in forensic mental health nursing. *Journal of Forensic Nursing*, 10, 234-242.
- Coid, J. W., Kallis, C., Doyle, M., Shaw, J., & Ullrich, S. (2015). Identifying causal risk factors for violence among discharged patients. *PLOS One*, 10(11), 1-17.
- de Vogel, V., Bruggeman, M., & Lancel, M. (2019). Gender-sensitive violence risk assessment: Predictive validity of six tools in female forensic psychiatric patients. *Criminal Justice and Behavior*, 46(4), 528-549.
- de Vogel, V. & de Vries Robbé, M. (2013). Working with Women. Towards a more gender sensitive violence risk assessment. In L. Johnstone & C. Logan (Eds.), *Managing Clinical Risk: A guide to effective practice* (pp. 224-241). London: Routledge.
- de Vogel, V., de Vries Robbé, M., Kalmthout, W.V., & Place, C. (2012). *Female Additional Manual (FAM). Additional guidelines to the HCR-20 for assessing risk for violence in women*. English version. Utrecht, The Netherlands: Van der Hoeven Kliniek.
- de Vogel, V., van den Broek, E., & de Vries Robbé, M. (2014). The use of the HCR-20 V3 in Dutch forensic psychiatric practice. *International Journal of Forensic Mental Health*, 13, 109-121.
- Douglas, K.S., & Belfrage, H. (2014). Interrater reliability and concurrent validity of the HCR-20 Version 3. *International Journal of Forensic Mental Health*, 13, 130-139.
- Douglas, K.S., Hart, S.D., Webster, C.D., & Belfrage, H. (2013). *HCR-20^{V3} Historical, Clinical, Risk Management (Version 3): Professional guidelines for evaluating risk of violence*. Burnaby, BC, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.
- Douglas, K.S., Webster, C.D., Hart, S.D., Eaves, D., & Ogloff, J.R.P. (2001). *HCR-20 Violence Risk Management Companion Guide*. Florida: Simon Fraser University.
- Doyle, M., Power, L.A., Coid, J., Kallis, C., Ullrich, S., & Shaw, J. (2014). Predicting post-discharge community violence in England and Wales using the HCR-20 V3. *International Journal of Forensic Mental Health*, 13, 140-147.
- Green, D., Schneider, M., Griswold, H., Belfi, B., Herrera, M., & DeBlasi, A. (2016). A comparison of the HCR-20V3 among male and female insanity acquittees: A retrospective file study. *International Journal of Forensic Mental Health*, 15(1), 48-64.
- Hogan, N.R., & Olver, M.E. (2016). Assessing risk for aggression in forensic psychiatric inpatients: An examination of five measures. *Law and Human Behavior*, 40(3), 233-243.
- Judges, R., Egan, V., & Broad, G. (2016). A critique of the Historical Clinical Risk-20, Version 3, risk assessment instrument. *Journal of Forensic Psychology Practice*, 16, 304-320.

Persson, M., Belfrage, H., Fredriksson, B., & Kristiansson, M. (2017). Violence during imprisonment, forensic psychiatric care, and probation: Correlations and predictive validity of the risk assessment instruments COVR, LSI-R, HCR-20^{V3}, and SAPROF. *International Journal of Forensic Mental Health*, 16(2), 117-129.

Smith, S.T., Kelley, S.E., Rulseh, A., Sorman, K., & Edens, J.F. (2014). Adapting the HCR-20V3 for pre-trial settings. *International Journal of Forensic Mental Health*, 13(2), 160-171.

Strub, D.S., Douglas, K.S., & Nicholls, T.L. (2014). The validity of Version 3 of the HCR-20 Violence Risk Assessment Scheme amongst offenders and civil psychiatric patients. *International Journal of Forensic Mental Health*, 13, 148-159.

Violence Risk Scale (VRS)

Daffern, M. (2007). The predictive validity and practical utility of structured schemes used to assess risk for aggression in psychiatric inpatient settings. *Aggression and Violent Behavior*, 12, 116-130.

Dolan, M., & Fullam, R. (2007). The validity of the Violence Risk Scale second edition (VRS-2) in a British forensic inpatient sample. *Journal of Forensic Psychiatry & Psychology*, 18, 381-393.

Dolan, M., Fullam, R., Logan, C., & Davies, G. (2008). The Violence Risk Scale Second Edition (VRS-2) as a predictor of institutional violence in a British forensic inpatient sample. *Psychiatry Research*, 158, 55-65.

Lewis, K., Olver, M.E., & Wong, S.C.P. (2012). The Violence Risk Scale: Predictive validity and linking changes in risk with violent recidivism in a sample of high-risk offenders with psychopathic traits. *Assessment*, 20, 150-164.

Ramesh, T., Igoumenou, A., Vazquez Montes, M., & Fazel, S. (2018). Use of risk assessment instruments to predict violence in forensic psychiatric hospitals: a systematic review and meta-analysis. *European Psychiatry*, 52, 47-53.

Wong, S., & Gordon, A. (1999). *Manual for the Violence Risk Scale*. Saskatchewan, Canada: University of Saskatchewan.

Wong, S.C.P., & Gordon, A. (2006). The validity and reliability of the Violence Risk Scale: A treatment-friendly violence risk assessment tool. *Psychology, Public Policy, and Law*, 12(3), 279-309.

Wong, S.C.P., & Parhar, K. (2011). Evaluation of the predictive validity of the Violence Risk Scale: A treatment-friendly violence risk assessment tool. *Psychology, Public Policy, and Law*, 22, 790-808.

Zhang, X., Chen, X., Cai, W., & Hu, J. (2012). Reliability of the Violence Risk Scale Chinese Version. *Journal of Forensic Medicine*, 28, 32-35.