NSW GOVERNMENT

Research Bulletin

Development and validation of the Triage Risk Assessment Scale: Sexual Offending (TRAS:SO)

Mark Howard, Yatin Mahajan, & Simon Corben

Aim

To develop and test an automated risk assessment tool that can be used to estimate sexual recidivism risk among men convicted of sexual offences, using routine administrative data assets maintained by Corrective Services NSW.

Method

Models were developed on a sample of 3630 men who were serving custodial orders in NSW associated with an index conviction for sexual offences. A series of bivariate tests of association and multivariable regression models were used to select predictors of sexual recidivism and develop predicted probability estimates for reoffending. In order to optimise cross-cultural validity, separate models were developed for Aboriginal and non-Aboriginal cohorts of men. Final models were validated using Receiver Operating Characteristic and other tests of predictive validity, in addition to bootstrapping techniques.

Results

The final predicted probability estimation models, which we have collectively named the TRAS:SO, showed strong discrimination accuracy for sexual recidivism within five years for the total sample (AUC = .830; 95% CI = .802-.857). Predictive validity for the Aboriginal cohort was satisfactory and significantly better than chance (AUC = .738; 95% CI = .666-.810), although lower than that for the non-Aboriginal cohort (AUC = .850; 95% CI = .821-.879). The TRAS:SO showed stronger discrimination performance compared to manual Static-99R assessments, among those men who had both assessments. Additional analyses indicated sound absolute predictive validity, as well as satisfactory stability of the models across simulated samples.

Conclusion

The results give promising indications that the TRAS:SO may be used to rapidly and accurately estimate sexual recidivism risk among men convicted of sex offences in NSW. The tool demonstrated predictive validity that is comparable to or greater than established benchmarks for manual assessment, while allowing for substantial time and resource savings that could potentially be reallocated to behaviour change intervention. Importantly, the TRAS:SO showed satisfactory performance for both Aboriginal and non-Aboriginal men, highlighting the potential value of such models in improving cross-cultural validity in local jurisdictions.

INTRODUCTION

Accurate assessment of sexual recidivism risk is a critical component in the effective case management of people convicted of sexual offences by corrections agencies. In alignment with principles of the Risk Need Responsivity (RNR: Bonta & Andrews, 2017) model, assessments of risk are necessary for core correctional decision-making such as allocation to and priority for behaviour change interventions. For several years, Corrective Services NSW has utilised the Static-99 and its variants (Hanson & Thornton, 2000; Helmus, Thornton et al., 2012) as routine risk assessments for people convicted of sexual offences. Currently in use by Corrective Services NSW, the Static-99R estimates sexual recidivism risk through calculation of a small number of historical variables that have empirically established associations with the likelihood of reoffending outcomes, such as age, sexual and non-sexual criminal history, and others. The Static-99R is perhaps the most commonly used assessment of sexual recidivism risk across jurisdictions, and has a robust history of validation and other research (see Helmus et al., 2022).

While the availability of tools such as the Static-99R are a positive foundation for risk assessment, there remain some challenges to their routine use by correctional agencies. A primary consideration is that manual assessments can be time and resource intensive, often requiring extensive work by highly qualified specialist staff such as psychologists. Resourcing implications become more pronounced when considered at the jurisdictional level. For example, at the end of 2022 there were 12,247 people in custody in the state of NSW, and around 20% of those serving a custodial sentence were recorded as having an index sex offence as their most serious offence (NSW Bureau of Crime Statistics and Research, 2022).

Another consideration is that the Static-99R and most other established risk assessment tools were not developed in Australia or with Australian samples, and there is often limited validation or normative data to support their ongoing use in local settings. A corollary is that there is limited evidence for cross-cultural validity for risk assessments in Australia, particularly in regards to culturally unique and diverse Aboriginal and Torres Strait Islander peoples (for the purposes of this report, the term 'Aboriginal' will hereafter be used to refer to all First Nations Australians, including Aboriginal and Torres Strait Islander peoples). This has raised criticism of various tools including the Static-99R in Australian clinical and legal contexts (Allan et al., 2018; 2020), which are compounded by growing evidence of poorer risk assessment performance for racial/ethnic minority groups and Indigenous peoples specifically (see Ahmed et al., 2023). While a small number of studies have examined the predictive validity of the Static-99 and Static-99R in Australia (Allan et al., 2006; Reeves et al., 2018; Smallbone & Rallings, 2013; Spiranovic, 2012), they have tended to be impacted by limitations such as small sample sizes.

One potential solution to the challenges of manual risk assessment is to develop automated actuarial estimates of risk using available archival or administrative data. Automation confers clear benefits to the time and personnel costs of risk assessment, and may also reduce opportunities for measurement error relative to manual methods. By training actuarial models on local data, there are also opportunities to improve and test the validity of assessment on local samples, including relevant cross-cultural groups. Given that established clinical assessments of sexual recidivism risk such as the Static-99R primarily generate estimates on the basis of criminal history and other static demographic variables, there appears to be potential to develop tools that perform similar functions through direct extraction and calculation from local criminal justice administrative databases.

Indeed, recent studies have indicated that automated tools can use small numbers of historical variables derived from Corrective Services NSW databases to generate risk estimates that have greater predictive validity compared to comprehensive manual assessments (e.g. Raudino et al., 2018; 2019; Howard et al., 2022).

Despite the potential advantages of automated assessment, a review of the literature indicates that there are few extant examples for assessing sexual recidivism risk, and we are aware of one cross-jurisdictional example currently employed in routine correctional operations. The Automated Sexual Recidivism (ASRS) was developed by the New Zealand Department of Corrections to facilitate their case management of men convicted of sexual offences under their jurisdiction (Skelton et al., 2006). It uses the New Zealand correctional database to automatically code seven of the ten measures of the Static-99. A subsequent revision, known as the ASRS-R, was generated to reflect changes to the age item weighting introduced by the Static-99R (Grace, 2014). Both the ASRS and ASRS-R have been found to have predictive validity outcomes that are comparable to manual assessments among people serving custodial orders for sexual convictions in New Zealand (Grace, 2014; Skelton et al., 2006).

Following this example, Corrective Services NSW has recently examined a number of potential models for generating automated estimates of sexual recidivism risk. Bell and Howard (2020) initially adopted the same approach as the ASRS and used local administrative databases to calculate scores for six items of the Static-99R, including age, index and prior non-sexual violence, prior sex offences, prior sentencing dates, and non-contact sex offences, for a sample of 3824 men serving custodial orders in NSW for sexual convictions. Model testing indicated relatively high levels of agreement between automated and manual scoring approaches to the items, and acceptable overall discrimination for sexual recidivism within 5 years (AUC = .68). A second iteration was also tested, which applied additional weightings to each item according to their estimated association with reoffending as derived from regression modelling. This approach was found to improve the predictive validity of estimates for sexual recidivism risk, relative to the unweighted model (AUC = .72).

The current study

Existing research has indicated that acceptably valid automated assessments of sexual recidivism risk can be generated by approximating items from the Static-99 and its variants (Bell & Howard, 2020; Grace, 2014; Skelton et al., 2006), and by reweighting those items to better estimate the relationships between predictor variables and outcomes in the local population (Bell & Howard, 2020). However, a potential limitation of these approaches is that they are bounded by existing variable selection and scoring rules established by the Static-99 or Static-99R, which may not translate well to data housed in local administrative databases or may omit other available variables that could contribute to predictive validity.

The aim of the current study is to expand on previous studies by developing an automated sexual recidivism risk assessment tool that adopts more comprehensive variable selection and modelling processes to optimise predictive validity within the context of local samples and data resources. To achieve this, we tested an unrestricted range of variables that may contribute to estimation of sexual recidivism risk and are routinely available through Corrective Services NSW databases, including but not limited to those previously employed in the Static-99R and other established sexual recidivism risk assessments. We also conducted a number of regression modelling and weighting techniques to more closely estimate associations between these variables and an

individual's likelihood of recidivism (Bell & Howard, 2000).

In recognition of evidence for relatively poor performance of sexual recidivism risk assessment tools for Indigenous peoples of Australia (Howard et al., 2023; Smallbone & Rallings, 2013; Spiranovic, 2012) and elsewhere (Ahmed et al., 2023), a related aim of this study was to develop a tool that accommodated cross-cultural validity considerations throughout and optimised predictive validity for Aboriginal men convicted of sexual offences. To this end we adopted modelling techniques that sought to detect and account for cultural differences across the development process, and conducted a number of analyses to test evidence for cross-cultural validity between Aboriginal and non-Aboriginal samples.

METHODS

Sample

The sample used in this study included adult men who were convicted of sex offences and had commenced a custodial episode with Corrective Services NSW from January 1999. In line with Static-99R eligibility criteria, men were required to have at least one conviction for 'category A' sex offences in their criminal history. We also considered only the first custodial episode for each individual, and any subsequent episodes were excluded to prevent model violations associated with non-independence of observations. Men in the sample were also required to have been released before 1 September 2017 to allow a minimum follow up period in the community of at least two years. These criteria resulted in a final sample of 3630 individuals for analysis. Among these, 663 men had an Aboriginal cultural background.

Data and measures

All data used for risk estimation models were retrieved from the Corrective Services NSW's Offender Information Management System (OIMS). OIMS is the central operational database maintained by Corrective Services NSW to manage people under supervision in custody and the community. It includes information on demographics, historical and current offences, results of assessments, and other case management and administrative processes.

A large pool of more than 50 potential predictor variables were extracted by OIMS for model development purposes. Variables were selected based on a rigorous review of available data from OIMS, and in reference to predictors identified in established sexual recidivism risk assessment tools and the broader sexual offending literature. Variables tended to cluster into those related to the individual's demographic characteristics, criminal history and sentencing characteristics, and history relating to sexual offending specifically. In the interests of brevity, the following gives an outline of variables that were selected for use in final predictive models:

- *Age:* calculated as the number of years between the individual's date of birth and the earliest possible release date. For the modelling, age was considered a continuous variable.
- *Total Corrective Services NSW episodes:* count of the total number of all custodial and community episodes prior to the current custodial episode. This total count was logarithmically transformed for modelling purposes and considered a continuous variable.
- *Prior violent offence:* a dichotomous variable indicating any non-sexual violent offence prior to the index offence (0 = No; 1 = Yes), included under Australian and New Zealand Standard Offence Classification (ANZSOC) subdivisions

0111, 0121, 0131, 0211, 0212, 0213, 0291, 0299, 0511, 0521, 0532, 0611, 0612, 0621, 1211, 1334, and 1531.

- *Prior non-sexual offences:* a dichotomous variable indicating the presence of a prior non-sexual offence and was calculated as the difference between the total number of prior convictions and the count of prior sex offences before the current index offence.
- *Prior community breaches:* a dichotomous variable indicating any breaches of community sentencing prior to the current custodial episode. This variable was considered a categorical variable (Yes/No).
- *Prior sexual offences:* total count of convictions for the following ANZSOC subdivisions attached to prior episodes: 0311, 0312, 0321, 0322, 0329, 0323, 1324, and 1325. This total count was logarithmically transformed for modelling purposes and considered a continuous variable.
- *Non-contact sexual offences:* a dichotomous variable indicating a prior or an index non-contact sexual offences classified under 0322, 0329 and 1325 ANZSOC subdivisions, including child pornography offences, non-assaultive sexual offence, and public order of sexual standards assault. This variable was considered a categorical variable (Yes/No).
- *Any division 0321 offence:* a dichotomous variable indicating a prior or an index non-assaultive sexual offence against a child prior to the current custodial episode. This variable was considered a categorical variable (Yes/No).
- *Prior division 0311 offence:* a dichotomous variable indicating a prior aggravated sexual assault before the current custodial episode. This variable was also considered a categorical variable (Yes/No).
- *Index division 0311 offence:* a dichotomous variable indicating an index aggravated sexual assault before the current custodial episode. This variable was also considered a categorical variable (Yes/No).
- *Prior division 1325 offence:* a dichotomous variable indicating a prior offence against public order of sexual standards assault before the current custodial episode. This variable was also considered a categorical variable (Yes/No).
- *Prior division 0322 offence:* a dichotomous variable indicating a prior child pornography offence before the current custodial episode. This variable was also considered a categorical variable (Yes/No).
- *Index division 0322 offence:* a dichotomous variable indicating an index child pornography offence before the current custodial episode. This variable was also considered a categorical variable (Yes/No).

The outcome variable was a dichotomous measure of whether or not men in the sample were recorded as being reconvicted in NSW criminal courts for a new sexual offence within five years following release from their index custodial episode. This data was derived from the NSW Bureau of Crime Statistics and Research (BOCSAR) Reoffending Database. Sexual reoffending was defined as a reconviction under the ANZSOC sexual assault and related offences division (03). Among all men in the sample who met the criteria for a minimum of 5 years follow-up period (n = 2713), 6.4% returned to CSNSW with a new sex offence conviction (n = 198) within five years of release.

Statistical analyses

Model development

A common series of steps was used to generate predictive models in this study. First, we examined bivariate relationships between prospective

predictor variables and the outcome variable as an initial selection method, using a series of binary logistic regression models. Variables were selected for further consideration if they had a significant bivariate relationship with sexual recidivism. Regression models were also used to compare deviance between different transformations or aggregates of variables to identify the best-fitting permutation. Correlations between predictor variables were also examined, and some variables were comparatively selected in the event that they had high shared variance.

Secondly, parameter estimates for sexual recidivism were examined after entering predictors into a multivariable regression model. In order to avoid overfitting to the model development sample, all predictor variables were entered simultaneously and sequential modelling was not used as a method of selection. The regression equation resulting from this model was then used to generate an estimate of each individual's probability of sexual recidivism.

Following other recent examples of risk assessment model development in local jurisdictions (Raudino et al., 2018; 2019; Howard et al., 2022), our initial approach was to apply the above steps to generate a single model for all men in the sample, including those of Aboriginal and non-Aboriginal background. Preliminary diagnostics indicated unsatisfactory performance of this model for Aboriginal men. Post-hoc methods of improving performance, including independent regression modelling of common predictor variables for each of the cohorts and other subgroup reweighting techniques, were found to have limited success.

We therefore opted to redevelop the tool by generating two distinct models for Aboriginal and non-Aboriginal cohorts. To achieve this, we reran bivariate variable selection techniques while examining consistencies or disparities in associations for Aboriginal and non-Aboriginal men, allowing for variability in the final set of items used in multivariable regression modelling for each of the groups. Consequently, regression models used to derive predictive probability estimates were conducted separately and included both different variables and different weighting coefficients between the groups. In this regard, the final TRAS:SO score as reported for the total sample comprised the predictive probability estimate derived from separate regression models for Aboriginal and for non-Aboriginal men.

Model validation

A number of techniques were used to test the validity and stability of TRAS:SO models, using an iterative process. In keeping with the aims of this study, tests were typically conducted for the Aboriginal and non-Aboriginal cohorts and models separately. Tests of TRAS:SO performance for the total sample were also conducted where appropriate.

Our primary indicator of predictive validity was relative accuracy, or performance in discriminating reoffenders from non-reoffenders. Discrimination performance was assessed using the Receiver Operating Characteristic (ROC) area under the curve (AUC) statistic. To interpret discrimination characteristics of risk assessments, AUC values of higher than .556 represent a small effect size; higher than .639 represent a medium effect size; and higher than .714 represent a large effect size (Rice & Harris, 2005). Model calibration was also assessed using the Hosmer-Lemeshow test, which indicates the goodness of fit or how well the observed sexual recidivism rates match the expected sexual recidivism rates across different score groups.

Discrimination performance was evaluated in reference to effect sizes described above, and also by comparison to available data for manual assessments of men in our sample using the Static-99R (Helmus et al., 2012). A total of n = 1066 men in the sample were found to have valid Static-99R assessments linked to their index custodial episode.

In the interests of comparability, AUC statistics for the TRAS:SO and Static-99R were recalculated and examined for only those individuals who had both assessments.

Finally, we conducted a validation of our two models by using bootstrapping to test the consistency of associations between predictor variables and recidivism across multiple samples for each model individually. Bootstrapping is a statistical technique that involves resampling the original dataset by drawing multiple random datasets. This process simulates sampling from an underlying population. Repeating the model estimation process on the multiple samples can then be used to examine if the variance in the model's regression coefficient across multiple random datasets of the same population is acceptable and if the model is stable. Bootstrapping results enable the replication of findings to be generalised without multiple samples of the original population dataset. In the current study, simulations were repeated 1000 times for both Aboriginal and non-Aboriginal cohorts by redrawing samples of the same size as the development sample with replacement, and repeating regression modelling for each of the samples. A key indicator of stability from the bootstrapping process is the resulting confidence intervals, which depict the range of regression coefficients (converted into odds ratios) derived from the simulated random samples (see Tables 1 and 2).

RESULTS

Initial model development processes indicated a number of predictor variables that were common to both Aboriginal and non-Aboriginal cohort models, including age; total Corrective Services NSW episodes; prior violent, non-sexual and sexual offences; prior breaches of community orders; prior or index non-contact sexual offences; and subgroups of sexual offence types including non-assaultive sexual offences against children (ANZSOC division 0321), aggravated sexual assault (division 0311), and offences against public order sexual standards (division 1325). For the non-Aboriginal cohort, additional variables were retained in the model relating to convictions for prior and index child pornography offences (division 0322). However, subsequent bootstrapping analyses showed signs of instability for particular variables marked by extreme confidence interval ranges, including the prior child pornography offences (division 0322) variable in the non-Aboriginal cohort model and offences against public order sexual standards (division 1325) variable in the Aboriginal cohort model[1]. These variables were therefore removed from their respective models before retesting and finalising the solution.

The results of logistic regression analyses for the final TRAS:SO models for Aboriginal and non-Aboriginal cohorts are given in the left panels of Tables 1 and 2. Key odds ratio (OR) statistics can be interpreted so that values of more than 1 indicate a positive association between the predictor and likelihood of recidivism, and values of 0-1 indicate a negative association between the predictor and likelihood of recidivism.

Results of logistic modelling for the Aboriginal cohort indicated that significant multivariate predictors of sexual recidivism included total count of CSNSW episodes and prior non-sexual violent offences. Men who had a greater number of CSNSW episodes were more likely to sexually reoffend. Conversely, men who had a prior violent offence were around a third as likely to sexually reoffend compared to those who did not (see Table 1).

[1] Model instability for these particular variables were likely attributable to low prevalence of convictions for these offences in the respective cohorts.

Table 1. Comparison of regression coefficients and error terms from the predictive model and averaged coefficient values and empirical distributions of error terms from the bootstrapped regression models for the Aboriginal sample.

Predictors	Predictive model					Bootstrapped model				
	B	SE	p	OR	[95% CI]	B	SE	p	OR	[95% CI]
Age at EPRD	.02	.01	.13	1.02	[.99, 1.05]	.02	.02	.15	1.02	[.98, 1.06]
Prior Sex offence count (log)	-.09	.43	.82	.90	[.39, 2.09]	-.09	.43	.79	.90	[.39, 2.10]
Total CSNSW episodes (log)	1.21	.28	<.001	3.37	[1.94, 5.84]	1.21	.30	<.001	3.37	[1.87, 6.07]
Prior Violent offence										
No	1					1				
Yes	-1.05	.45	.02	.35	[.14, .84]	-1.05	.50	.02	.35	[.13, .93]
Prior community breaches										
No	1					1				
Yes	.05	.36	.89	1.05	[.51, 2.15]	.05	.37	.88	1.05	[.51, 2.17]
Prior non-sex offences										
No	1					1				
Yes	-.18	.53	.73	.84	[.30, 2.35]	-.18	.57	.70	.84	[.27, 2.55]
Prior non-contact offences										
No	1					1				
Yes	.52	.39	.17	1.69	[.79, 3.61]	.52	.39	.15	1.69	[.79, 3.61]
Any 0321 (non-assaultive sexual offences against a child)										
No	1					1				
Yes	.32	.60	.60	1.38	[.43, 4.50]	.32	1.40	.56	1.38	[.09, 21.5]
Prior 0311 (Aggravated sexual assault)										
No	1					1				
Yes	.57	.60	.33	1.78	[.55, 5.77]	.57	.64	.31	1.78	[.51, 6.25]
Index 0311 (Aggravated sexual assault)										
No	1					1				
Yes	.08	.38	.82	1.10	[.52, 2.28]	.08	.39	.82	1.10	[.52, 2.41]

For non-Aboriginal men, examination of general, criminal history-related, and sexual-offences-related variables entered into the logistic regression model indicated that sexual recidivism was significantly associated with having a greater number of CSNSW episodes; history of a violent offence; an indication of any non-assaultive sexual offences against a child; having an index aggravated sexual assault and having an index child pornography offence. For instance, men with an index child pornography offence were more than three times as likely to be reconvicted for sexual reoffending within five years of their release compared to men with no index child pornography offence (see Table 2).

Table 2. Comparison of regression coefficients and error terms from the predictive model and averaged coefficient values and empirical distributions of error terms from the bootstrapped regression models for the non-Aboriginal sample.

Predictors	Predictive model					Bootstrapped model				
	B	SE	p	OR	[95% CI]	B	SE	p	OR	[95% CI]
Age at EPRD	-.01	.01	.40	.99	[.98, 1.00]	-.01	.01	.40	.99	[.98, 1.01]
Prior Sex offence count (log)	.27	.20	.17	1.32	[.89, 1.96]	.27	.20	.14	1.32	[.89, 1.96]
Total CSNSW episodes (log)	1.6	.17	<.001	4.96	[3.55, 6.90]	1.6	.18	<.001	4.96	[3.48, 7.04]
Prior Violent offence										
No	1					1				
Yes	-.75	.28	.006	.47	[.27, .80]	-.75	.29	.01	.47	[.27, .83]
Prior community breaches										
No	1					1				
Yes	-.37	.25	.12	.69	[.42, 1.12]	-.37	.29	.19	.69	[.39, 1.21]
Prior non-sex offences										
No	1					1				
Yes	.16	.28	.54	1.18	[.69, 2.02]	.16	.30	.57	1.18	[.66, 2.13]
Prior non-contact offences										
No	1					1				
Yes	.35	.25	.16	1.40	[.87, 2.32]	.35	.27	.17	1.4	[.84, 2.41]
Any 0321 (non-assaultive sexual offences against a child)										
No	1					1				
Yes	.53	.25	.03	1.70	[1.04, 2.80]	.53	.27	.04	1.70	[1.00, 2.89]
Prior 0311 (Aggravated sexual assault)										
No	1					1				
Yes	.11	.38	.76	1.12	[.53, 2.36]	.11	.39	.75	1.12	[.52, 2.41]
Index 0311 (Aggravated sexual assault)										
No	1					1				
Yes	-.42	.22	.06	.66	[.43, 1.00]	-.42	.23	.07	.66	[.42, 1.03]
Index 0322 (Child pornography offences)										
No	1					1				
Yes	1.15	.36	.001	3.17	[1.58, 6.38]	1.15	.39	.003	3.17	[1.48, 6.82]
Prior 1325 (Offences against public order sexual standards)										
No	1					1				
Yes	.49	.42	.24	1.64	[.72, 3.77]	.49	.42	.22	1.64	[.72, 3.77]

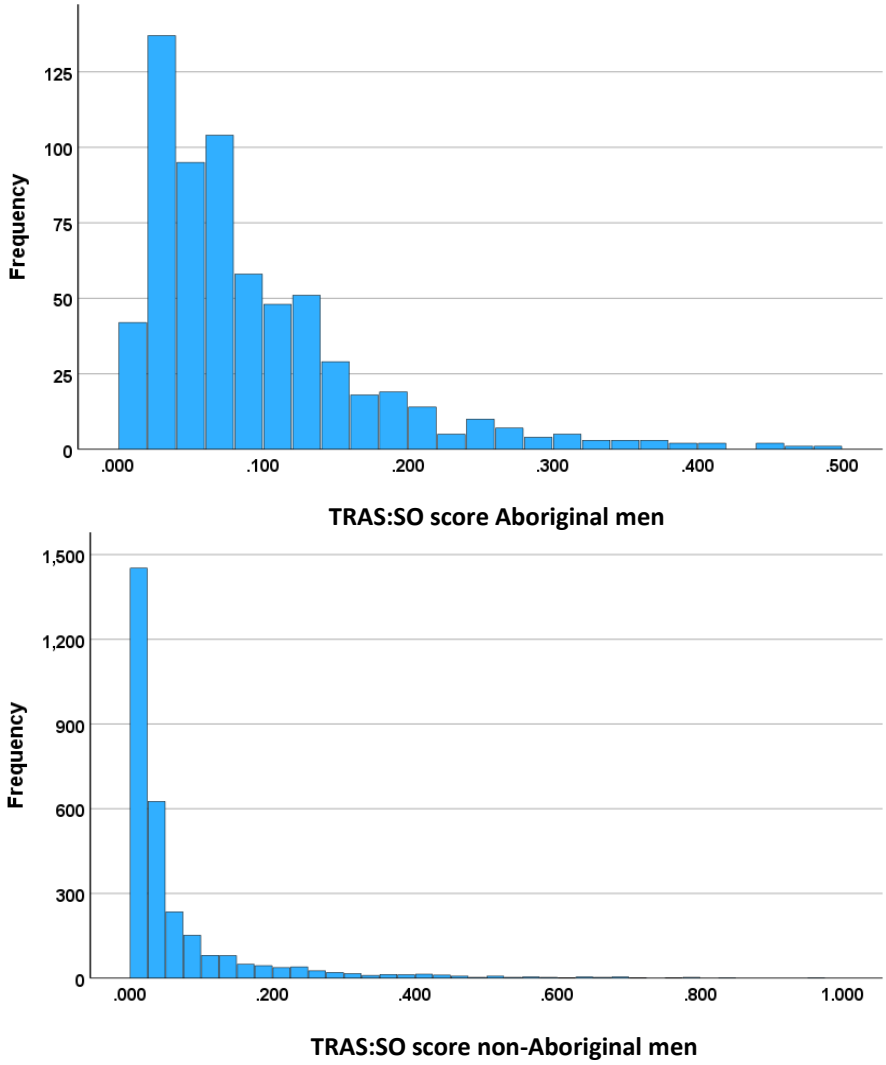

Figure 1. Distribution of TRAS:SO probability estimates for Aboriginal men (top panel) and for non-Aboriginal men (bottom panel)

Figure 1 shows the distribution of TRAS:SO scores for the Aboriginal and non-Aboriginal cohorts of men. The mean predicted probability of sexual recidivism for the non-Aboriginal sample, as estimated by TRAS:SO, was .06 (SD = .10) with a range of .01 - .97. For Aboriginal men, the mean predicted probability of sexual recidivism was .09 (SD = .08) with a range of .01 - .47. TRAS:SO scores can be interpreted so that on average, non-Aboriginal and Aboriginal men had 6% and 8% likelihood of sexual recidivism within 5 years of release from their index custodial episode, respectively.

Model validation

Discrimination performance of the TRAS:SO was assessed against the criterion of sexual recidivism within 5 years of release, and relevant AUC statistics are given in Table 3. The TRAS:SO was found to return AUC values of .738 for Aboriginal men and .850 for non-Aboriginal men. After aggregating estimates for both cohorts, TRAS:SO discrimination performance for the total sample corresponded to an AUC value of .830. These AUC values correspond to large effect sizes in discriminating reoffenders from non-reoffenders based on TRAS: SO scores.

Table 3. AUC statistics and 95% confidence intervals for estimates of sexual recidivism within five years derived from the TRAS:SO and STATIC-99R.

Sample	n	TRAS: SO		STATIC-99R	
		AUC	[95% CI]	AUC	[95% CI]
Total Sample					
All men	3630	.830	[.802 - .857]	--	--
Aboriginal cohort	663	.738	[.666 - .810]	--	--
Non-Aboriginal cohort	2967	.850	[.821 - .879]	--	--
STATIC-99R Sample					
All men	1064	.863	[.812 - .915]	.805	[.735 - .876]

Discrimination accuracy was also examined by comparing performance between the TRAS:SO and manual Static-99R assessments, for those men in the sample who had received a Static-99R in their index custodial episode (n = 1064; see Table 3). Due to low cell sizes for Aboriginal men, only AUC statistics for the full Static-99R sample were assessed. It can be seen that among this group of men, the TRAS:SO returned an AUC value of .863 for sexual recidivism within 5 years. This compares to an AUC value of .805 for the Static-99R.

To test model calibration, we examined the Hosmer-Lemeshow test statistic derived from final regression models. For both non-Aboriginal ($\chi 2$ (8) = 13.17, p = .11) and Aboriginal ($\chi 2$ (8) = 2.66, p = .95) cohorts, the test values were not statistically significant. These results indicate non-significant differences between observed and estimated rates of sexual recidivism for both cohorts within each score group.

Model Stability

To test the stability of the TRAS:SO, we used bootstrapping as the primary validation process. The bootstrapping results for both Aboriginal and non-Aboriginal cohort final models showed that the coefficients and error distribution estimates were similar to those obtained from the original regression models (see Tables 1 & 2). In particular, the bootstrapped 95% confidence intervals had small and comparable ranges for predictors across both models. The results indicate a good degree of reliability, or reproducibility of coefficient values across simulated samples, for both of the regression models.

DISCUSSION

The aim of this study was to develop an actuarial tool that can rapidly and accurately generate automated estimates of sexual recidivism risk among men convicted of sexual offences, which we have named the TRAS:SO. Our results indicated that the TRAS:SO had good predictive validity for the target population. As a benchmark index of discrimination performance, the AUC value for sexual reoffending within five years was .830 for the total sample. This is substantially above average relative to available statistics on manual assessments; for example, an international meta-analysis of 56 studies of the Static-99R returned average AUC values of .68-.69 at a follow up of five years (Helmus et al., 2022), while a recent study of Static-99R outcomes among men in NSW returned an AUC value of .76 (Howard et al., 2023). Consistent with this, we found that the TRAS:SO showed stronger relative predictive validity compared to the Static-99R for men in our sample who had both assessments attached to their index custodial episode. Additional analyses also showed

promising indications of calibration between observed and expected recidivism rates, as well as model stability across Aboriginal and non-Aboriginal cohorts of men.

It is also noted that the TRAS:SO showed discrimination performance that improved upon previous automated sexual recidivism risk assessment models developed by Corrective Services NSW using similar samples (Bell & Howard, 2020). The two primary innovations of the TRAS:SO relative to previous models included consideration of a range of available variables as opposed to those that corresponded with Static-99R items, and more extensive statistical modelling to account for differences between and optimise predictive validity for Aboriginal and non-Aboriginal cohorts of men. These factors emphasise some of the potential benefits of automated risk assessment on performance outcomes, including capabilities to tailor models to local data sources and samples, and opportunities for more complex calculations of predictor variables and their relationships with outcomes compared to what might be feasible through manual assessment processes.

In line with our aims to improve upon cross-cultural validity as part of model building and testing processes, the TRAS:SO showed sound discrimination performance for Aboriginal men (AUC = .738) that was significantly better than chance. While small sample sizes precluded meaningful comparisons of performance between the TRAS:SO and Static-99R for Aboriginal men in this study, recent research on the Static-99R observed AUC values of .682 for sexual recidivism within five years among Aboriginal men in NSW (Howard et al., 2023), which in turn was higher than that recorded for other international Indigenous samples on average (e.g., Ahmed et al., 2023). We are aware of one previous study that has reported higher discrimination values for Aboriginal Australian men convicted of sexual offences, whereby Smallbone & Rallings (2013) found an AUC value of .76 for the original Static-99; however, it is noted that this result was observed with a substantially smaller cohort sample (n = 67) and shorter follow up times (average = 29 months). Evidence for the performance of the TRAS:SO with Aboriginal men gives promising indications for the potential value of locally developed actuarial models for addressing important concerns about poor (or unknown) predictive validity of risk assessment tools for culturally diverse peoples.

While this study indicated that the TRAS:SO may improve on other methods of assessing sexual recidivism risk among Aboriginal men, it is nonetheless noted that discrimination performance for this cohort was markedly lower than for non-Aboriginal men. This follows an international pattern of relative underperformance of risk assessments for Indigenous peoples (e.g. Ahmed et al., 2023). Such a disparity could be indicative of multiple forms of cultural bias in risk assessment, including in how risk assessments are administered, relationships between given risk factors and recidivism outcomes, and differences in contributors to detected recidivism such as overpolicing (Ahmed et al., 2023; Helmus et al., 2011; Wilson & Gutierrez, 2014). Another potential contributing factor relates to statistical artefacts arising from differences in sample composition across groups. For example, Howard and colleagues (2023) found that differences in Static-99R predictive validity for Aboriginal and Non-Aboriginal men in NSW were attenuated after matching cohorts on their risk profiles, and attributed this to higher average scores and restriction in range effects on AUC values for the Aboriginal cohort (see also Helmus et al., 2022). Consistent with this, we found that the TRAS:SO tended to generate a smaller range of estimates for Aboriginal men relative to non-Aboriginal men. In any case, it is important that such disparities are subject to ongoing scrutiny in future research to better understand causal mechanisms and improve upon cross-cultural validity of risk assessments.

Limitations

Some limitations of the study are noted. Following from the observations above, while this study included almost all men serving custodial orders for sexual offences in NSW over a period of almost 20 years, the available sample for Aboriginal men was nonetheless suboptimal. Smaller samples are likely to influence the stability of model coefficients and can also introduce compromises in model development; for example, exclusions of predictor variables in the TRAS:SO models often appeared to be related to small numbers of observed events (such as a given type of conviction) within the sample. It is possible that such factors could have contributed to a lower range of estimated TRAS:SO values for the smaller group of Aboriginal men relative to non-Aboriginal men, with lead-on effects for primary AUC statistics used for measuring discrimination performance. Importantly, we also acknowledge that the TRAS:SO was modelled with men serving custodial orders in NSW, who may not be representative of the rich cultural diversity of First Nations peoples across Australia.

Other limitations relate to the quality of the data used to develop and test the TRAS:SO. Both predictor variables and recidivism outcomes were derived from administrative databases that pertain to criminal justice activities recorded in the state of NSW only. As a result, the predictive accuracy of historical variables used in the model, as well as detection of recidivism, may be impacted in cases of interjurisdictional movements and offending. A related factor is that while automation processes such as those used by the TRAS:SO could serve to reduce error associated with manual calculation of assessment items, their utility is nonetheless affected by the quality and stability of underlying data records and associated entry processes. In this regard, the TRAS:SO would likely benefit from periodic remodelling to account for changes in variable definitions or recording processes in administrative databases over time.

Conclusions

The results of this study give promising indications that the TRAS:SO can be used to quickly and accurately estimate risk of sexual recidivism among men convicted of sexual offences in NSW. It is consistent with the largely historical and static nature of variables used in established risk assessments for sexual recidivism, such as the Static-99R, that effective tools for estimating likelihood of reoffending may be derived by leveraging relevant information stored in administrative databases. To this end, our analyses indicated that the TRAS:SO achieved similar or better discrimination performance than the Static-99R, while substantially reducing the time and other resource costs involved in its administration. Importantly, the TRAS:SO showed evidence of sound predictive validity for both Aboriginal and non-Aboriginal men, highlighting the potential value of developing actuarial models for assessing risk within local jurisdictions.

Because the TRAS:SO only utilised standardised variables that are routinely available in Corrective Services NSW operational databases, it has the potential to be automated within existing data systems to generate almost instantaneous estimates of sexual recidivism risk for large cohorts of individuals. The efficiency gains afforded by this, and other automated risk assessment models, may in turn allow for increased allocation of the limited specialist clinical resources available to corrections agencies towards behaviour change interventions that seek to ameliorate risk of recidivism among men convicted of sexual offences.

REFERENCES

Ahmed, S., Lee, S. C., & Helmus, L. M. (2023). Predictive accuracy of Static-99R across different racial/ethnic groups: A meta-analysis. *Law and Human Behavior, 47*(1), 275-291. https://doi.org/10.1037/lhb0000517

Allan, A., Dawson, D., & Allan, M.M. (2006). Prediction of the risk of male sexual reoffending in Australia. *Australian Psychologist, 41*(1), 60–68. doi:10.1080/00050060500391886

Allan, A., Parry, C. L., Ferrante, A., Gillies, C., Griffiths, C. S., Morgan, F., & Wong, S. C. P. (2018). Assessing the risk of Australian Indigenous sexual offenders reoffending: A review of the research literature and court decisions. *Psychiatry, Psychology and Law, 26*(2), 274–294. https://doi.org/10.1080/13218719.2018.1504242

Allan, A., Parry, C.L., Tubex, H., Spiranovic, C., & Morgan, F. (2020). Forensic experts' perspectives on Australian Indigenous sexual offenders and factors important in evaluating the risk of recidivism. *International Journal of Offender Therapy and Comparative Criminology*, 65(15), 1653–1675. https://doi.org/10.1177/0306624X20967941

Bell, A., & Howard, M. (2020). *Automated assessment of sexual recidivism risk for custody-based sex offenders*. Sydney: Corrective Services NSW.

Bureau of Crime Statistics and Research (2023). *New South Wales Custody Statistics, Quarterly update December 2022.* Sydney: NSW Bureau of Crime Statistics and Research.

Bonta, J., & Andrews, D. A. (2017). *The psychology of criminal conduct (6th ed.).* Routledge.

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24*(1), 119–136. https://doi.org/10.1023/A:1005482921333

Helmus, L. M., Kelley, S. M., Frazier, A., Fernandez, Y. M., Lee, S. C., Rettenberger, M., & Boccaccini, M. T. (2022). Static-99R: Strengths, limitations, predictive accuracy meta-analysis, and legal admissibility review. *Psychology, Public Policy, and Law, 28*(3), 307–331. https://doi.org/10.1037/law0000351

Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse: A Journal of Research and Treatment 24*(1), 64–101. https://doi.org/10.1177/1079063211409951

Howard, M.V.A., Chong, C.S., & Murphy, K. (2023). *Static-99R norms and cross-cultural validity for Australian Aboriginal and non-Aboriginal men convicted of sexual offences* [Manuscript submitted for publication]. Corrective Services NSW.

Howard, M., Wei, Z., Zhang, Y., & Corben, S. (2022). *Actuarial assessment of domestic violence recidivism risk among custody-based males: The Domestic Violence – Triage Risk Assessment Scale (DV-TRAS).* Sydney: Corrective Services NSW.

Grace, R.C. (2014). *The Automated Sexual Recidivism Scale (ASRS): Revised age weights.* Unpublished manuscript commissioned by NZ Department of Corrections.

Raudino, A., Corben S., Galouzis, J., Mahajan, Y., & Howard, M. (2019). *The Custody Triage Risk Assessment Scale (Custody TRAS): An updated statistical model for predicting risk of return to custody.* Sydney: Corrective Services NSW.

Raudino, A., Corben S., van Doorn, G & Galouzis, J. (2018). *The Community Triage Risk Assessment Scale (Community TRAS): A statistical model for predicting recidivism among community-based offenders.* Sydney: Corrective Services NSW.

Reeves, S.G., Ogloff, J.R.P., & Simmons, M. (2018). The predictive validity of the Static-99, Static-99R, and Static-2002/R: Which one to use? *Sexual Abuse: A Journal of Research and Treatment, 30*(8), 887–907. doi:10.1177/1079063217712216

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, 29*, 615–620. https://doi.org/10.1007/s10979-005-6832-7

Skelton, A., Riley, D., Wales, D., & Vess, J. (2006). Assessing risk for sexual offenders in New Zealand: Development and validation of a computer-scored risk measure. *Journal of Sexual Aggression, 12*(3), 277–286. https://doi.org/10.1080/13552600601100326

Smallbone, S., & Rallings, M. (2013). Short-term predictive validity of the Static-99 and Static-99R for indigenous and nonindigenous Australian sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 25*(3), 302–316. doi:10.1177/1079063212472937

Spiranovic, C. (2012). *The Static-99 and Static-99-R Norms Project: Developing Norms Based on Western Australian Sex Offenders.* Crawley, Western Australia: University of Western Australia.

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*(2), 196–219. https://doi.org/10.1177/0093854813500958

Other CRES Research Titles

July 2023	Implementation of an initiative for people with less than five months to serve in prison: Staff insights on the Short Sentence Intensive Program (SSIP)
June 2023	The Intensive Learning Centre: Staff experiences of implementation and the SILC reforms
April 2023	Five Minute Interventions (FMI): Custodial staff views and experiences of implementing FMI
Dec 2022	Effects of the Workplace Mentor Program on correctional officers' perceptions of workplace culture
Dec 2022	Five Minute Interventions (FMI): Long-term effects of training on custodial staff attitudes towards prisoners, motivation and ability to support rehabilitation, and job stress and satisfaction
Nov 2022	Effects of Words @ Work training on perceptions of offender rehabilitation and job experiences among Corrective Services Industries (CSI) overseers
Sept 2022	Offender participation outcomes and predictors of treatment completion in the High Intensity Program Units (HIPUs)
Aug 2022	Staff and inmate experiences of prison social climate at Rapid Build correctional centres: A quantitative evaluation
July 2022	Evaluation of the Alternate Sanctions Program (ASP): Within-treatment Change
June 2022	Implementing digital technologies in prisons: Inmate uptake and perceived value of in-cell digital tablets
April 2022	Actuarial assessment of domestic violence recidivism risk among custody-based males: The Domestic Violence – Triage Risk Assessment Scale (DV–TRAS)
March 2022	A qualitative exploration of factors influencing prison social climate at Rapid Build and traditional correctional centres
Jan 2022	Quality of the therapeutic alliance and associations with program outcomes for offenders participating in High Intensity Program Units (HIPUs)
Dec 2021	The Initial Transitional Support (ITS) program: Implementation evaluation
Dec 2021	How does the role of custodial staff influence their perceptions of offender rehabilitation and responses to Five Minute Interventions (FMI) training?
Oct 2021	Five Minute Interventions (FMI): Short-term effects of training on staff attitudes towards prisoners, motivation and ability to support rehabilitation, and job stress and satisfaction
Sept 2021	Process evaluation of the Custody Based Intensive Treatment (CUBIT) program for sex-offenders: Within-treatment change
Sept 2021	Impact Evaluation of the Gurnang Life Challenge Specialised Program for Young Adult Male Offenders in NSW
June 2021	Evaluation of Custodial Case Management (CCM): The Planning for Adjustment, Responsivity, Reintegration, Criminogenic Needs, and Communication (PARRCC) Assessment Tool
March 2021	Evaluation of High Intensity Programs Units (HIPUs): Implementation of an innovative intervention model for offenders with short custodial sentences
March 2021	Women in prison: An examination of the support needs of women in custody with children
Feb 2021	The Initial Transitional Support (ITS) program: A profile of offender participation and service delivery
Oct 2020	Automated assessment of sexual recidivism risk for custody-based sex offenders
Oct 2020	The predictive validity of general risk assessment tools for offence-specific recidivism among domestic violence offenders

Research Bulletin No.60
ISSN 2207 0850

Corrections Research, Evaluation & Statistics
Governance & Continuous Improvement
Corrective Services NSW
GPO Box 31
Sydney NSW Australia

Telephone: (02) 8346 1556
Email: research.enquiries@justice.nsw.gov.au