

## The predictive validity of general risk assessment tools for offence-specific recidivism among domestic violence offenders

Mark Howard and Yun Zhang

**AIM** To examine whether risk assessment tools for general recidivism, including the Level of Service Inventory - Revised (LSI-R) and the Triage Risk Assessment Scale (TRAS) measures developed by Corrective Services NSW, have predictive validity for domestic violence (DV) recidivism.

### **FINDINGS AND CONCLUSIONS**

For offenders in custody, both the LSI-R and Custody TRAS had acceptable predictive validity for general recidivism among all offenders in the sample. Predictive accuracy showed incremental declines when assessing general recidivism and again when assessing DV recidivism among DV offenders. Discrimination performance of the LSI-R (AUC = .59) and the Custody TRAS (AUC = .59) was poor when assessing DV-related recidivism among custody-based offenders, and was only marginally better than chance in some cases.

For offenders in the community, the LSI-R and Community TRAS also had acceptable predictive validity for general recidivism. Incremental declines when assessing DV recidivism were also observed; however these were less pronounced than for offenders in custody. Community-based assessments using the LSI-R (AUC = .67) and the Community TRAS (AUC = .69) showed fair performance in discriminating DV recidivism.

Follow up analyses showed that assessments had poorer specificity, or were more likely to inaccurately assess a non-recidivist as having high risk, for DV recidivism than for general recidivism. Differences in definitions of recidivism also contributed to variation in discrimination performance.

We concluded that while general risk assessment tools can have predictive validity for offence-specific recidivism among DV offenders, the degree of variability in performance across samples and outcomes precluded their consistent use as a primary measure of DV reoffending risk. However, they may have promise as a means of screening out DV offenders who are at lower risk of serious DV-related recidivism and triaging higher risk offenders to offence-specific assessments.

## INTRODUCTION

---

Domestic violence (DV) is a prevalent public health issue in Australia which causes significant social, emotional, and economic costs to victims, their families and the broader community (Carrington & Phillips, 2006; Cox, 2015; Laing & Bobic, 2002). Growing recognition of this issue over recent years has been accompanied by increasing use of criminal justice and other social interventions to reduce DV-related offending and risks of future reoffending. In New South Wales (NSW) this is exemplified by the recent Premier's Priority to reduce DV reoffending by 5% over a five year period.

Corrective Services NSW delivers a number of interventions targeted at reducing risk of reoffending among DV offenders who are under supervision in custody and the community. Foremost among these is a series of group behaviour change programs known as EQUIPS. The EQUIPS Domestic Abuse program is widely available to eligible DV offenders and addresses specific needs associated with violence towards intimate partners. Given that DV offenders often have a range of criminogenic needs for both DV and other offending (e.g. Zhang & Howard, 2020), they can also attend multiple modules of EQUIPS interventions to address other violent behaviours (EQUIPS Aggression), substance use (EQUIPS Addiction) and general risk factors for reoffending (EQUIPS Foundation) in addition to EQUIPS Domestic Abuse.

Following the risk principle of the risk-need-responsivity (RNR) model of correctional intervention (Andrews & Bonta, 2010), for programs such as EQUIPS to be effective they should be delivered to offenders who are at higher risk of reoffending. This has a relationship to the needs principle, where more intensive intervention is required to address more severe needs that are causally related to the target offending behaviours. In the case of offence-specific programs such as EQUIPS Domestic Abuse, it follows that intervention to address DV-related risk factors should be delivered to offenders who have higher risk of DV reoffending in particular, as opposed to higher risk of reoffending in general.

Adherence to the risk principle requires accurate assessment of offenders' likelihood of reoffending. In many jurisdictions, offender case management and allocation to various interventions is guided by assessment of general risk of recidivism, using validated tools such as the Level of Service Inventory - Revised (LSI-R: Andrews & Bonta, 1995). While likelihood of general recidivism is a valid index of suitability for interventions to address common risk factors for reoffending, it is not necessarily the case that such measures can be generalised to risk for particular categories of offending and suitability for offence-specific interventions. For example, it has been established that future sexual offending is often unrelated to assessed risk of general reoffending (e.g. Ragusa-Salerno, Ostermann, & Thomas, 2013), leading to the common use of offence-specific risk assessments to guide decision making about intervention strategies for sex offenders.

A large number of studies have examined the predictive validity of general risk assessment tools such as variants of the LSI-R (for a review see Olver, Stockdale, & Wormith, 2014), and there are indications that these tools can have utility in discriminating risk for offence categories such as violent reoffending. There is less evidence for the predictive validity of general risk assessments for DV-specific reoffending. Studies have found that variants of the LSI-R have significant although modest accuracy in predicting DV outcomes (Hanson, Helmus, & Bourgon, 2007; Hendricks, Werner, Shipway, & Turinetti, 2006; Hilton, Harris, & Rice, 2010), or do not predict DV reoffending significantly better than chance (Hilton, Harris, Popham, & Lang, 2010).

The aim of this study is to examine whether tools that are routinely used by Corrective Services NSW to assess general recidivism risk have predictive validity for likelihood of DV-related reoffending, among offenders who have recent convictions for DV offending. Tools assessed in this study include the LSI-R as well as two actuarial measures that were developed by Corrective Services NSW to support risk assessment and case management decisions at the offender population level, known as the Custody Triage Risk Assessment Scale (Custody TRAS) and the Community TRAS.

The Custody TRAS uses various demographic and criminal history characteristics (including age, previous convictions, density of custodial episodes, and others) to estimate probability of return to custody within two years, among offenders serving custodial sentences (Raudino, Corben, Galouzis, Mahajan, & Howard, 2019). The Community TRAS similarly uses static variables to estimate probability of return to any corrections episode within two years, among offenders serving community-based orders (Raudino, Corben, van Doorn, & Galouzis, 2018).

It is intended that this study will help to support decision making about the utility of risk assessment approaches to inform allocation of DV offenders to offence-specific interventions.

## METHODS

---

Sampling for this study was based on samples of offenders who were included in initial validation projects for the Custody TRAS and the Community TRAS. Offenders in the Custody TRAS sample included all adults who were released from full-time custody in NSW between January 2011 and December 2014. Offenders in the Community TRAS sample included all those who entered community supervision between July 2010 and June 2013 (for further information about the samples, see Raudino et al., 2018; 2019). To be included in the current study, offenders in the Custody TRAS ( $n = 17098$ ) and Community TRAS ( $n = 24919$ ) samples were also required to have been administered a valid LSI-R assessment during their index episode.

Target samples of DV offenders were also derived from the broader Custody TRAS and Community TRAS samples for the purposes of this study. DV offenders were identified as those who had any conviction with a DV Lawpart code attached to their index corrections episode. There were 3601 DV offenders in the Custody TRAS sample and 5437 DV offenders in the Community TRAS sample who were eligible for the current study.

Recidivism outcomes used the same definitions as those employed for development of the Custody TRAS and Community TRAS, and differed for custody and community samples. For custody-based offenders, general recidivism was measured as any return to custody with a new sentence (with or without balance of parole) within two years. For community-based offenders, general recidivism was measured as new convictions resulting in a custodial or supervised community sentence within two years. DV recidivism outcomes used similar definitions with the specifier that new sentences included convictions for DV-related offending, as indicated by a DV Lawpart code.

Primary analyses in this study tested the predictive validity of assessments using receiver operating characteristic (ROC) area under the curve (AUC) statistics. AUC statistics are a widely used measure of discrimination accuracy which assesses the probability that a score randomly selected from a group with the signal of interest (in this instance, offenders who recidivate) is higher than a score selected from a

group without the signal of interest (in this instance, offenders who do not recidivate). AUC values can be interpreted so that .50 indicates discrimination that is no better than chance and 1.0 indicates perfect discrimination. Hosmer and Lemeshow (2000) suggest rules of thumb to classify discrimination performance so that AUC values of .50 to .60 are ‘poor’, .60 to .70 are ‘fair’, .70 to .80 are ‘acceptable’, .80 to .90 are ‘excellent’ and .90 or above are ‘outstanding’.

Analyses were conducted using continuous scores generated by the LSI-R (range = 0 – 54), the Custody TRAS (range = .02 – .89) and the Community TRAS (range = .06 – 1.00), unless otherwise noted.

## FINDINGS

### Risk and reoffending characteristics of DV offenders

Table 1 shows rates of general and DV recidivism among DV offenders in the custody and community samples, in addition to descriptive statistics for applicable risk assessments. As a point of comparison, we also include recidivism and risk assessment outcomes for other offenders in the samples who did not have index DV offences.

For offenders in the custody sample, it can be seen that DV offenders had slightly higher risk of general recidivism, as assessed by the LSI-R and the Custody TRAS, compared to non-DV offenders. DV offenders also had higher rates of general recidivism within two years than non-DV offenders. Almost one in five (18.2%) DV offenders returned to custody for DV-related recidivism within two years, which was substantially higher than non-DV offenders (4.5%).

For offenders in the community sample, DV offenders and non-DV offenders had similar assessed risk of general recidivism on average. Similar to the custody sample, community-based DV offenders had higher rates of general recidivism within two years and were around four times as likely to be reconvicted for DV-related recidivism, compared to non-DV offenders.

**Table 1.** Descriptive statistics (M(SD) or %) for average risk assessment scores and recidivism outcomes among DV and non-DV offenders in the custody and community samples.

Measure	Custody sample			Community sample		
	DV	Non-DV	Total	DV	Non-DV	Total
TRAS score	.30 (.16)	.28 (.18)	.28 (.17)	.43 (.21)	.42 (.21)	.42 (.21)
LSI-R total score	31.06 (7.45)	27.91 (9.51)	28.60 (9.20)	22.54 (8.71)	20.77 (9.23)	21.2 (9.15)
General recidivism (%)	47.5%	38.7%	40.5%	43.7%	39.8%	40.6%
DV recidivism (%)	18.2%	4.5%	7.4%	25.3%	6.5%	10.6%

Note: TRAS refers to Custody TRAS for the custody sample and Community TRAS for the community sample.

### Predictive validity of risk assessments

Table 2 gives AUC statistics assessing predictive validity of the LSI-R and TRAS tools for DV and general recidivism, among offenders with a DV-related index offence. Predictive validity statistics for general recidivism among all offenders in the custody and community samples are also given, which serves as a benchmark for discrimination when the tools are used for their original purpose.

The Custody TRAS and LSI-R showed acceptable discrimination for general recidivism when assessed for all offenders in the custody sample. This declined slightly when assessing general recidivism outcomes for custody-based DV offenders only. AUC statistics indicated markedly poorer performance in predicting DV recidivism outcomes for DV offenders, with both the Custody TRAS and LSI-R returning values of .59. Confidence intervals did not include .5, indicating that the discrimination performance of both tools for DV recidivism was significantly better than chance.

**Table 2.** AUC statistics indicating discrimination accuracy of risk assessment tools for general recidivism among all offenders, and for general and DV recidivism among DV offenders, in the custody and community samples.

Sample	Recidivism	Custody / Community TRAS		LSI-R	
		AUC	[95% CI]	AUC	[95% CI]
Custody					
All offenders	General	.76	[.75-.77]	.71	[.71-.72]
DV offenders	General	.69	[.67-.71]	.66	[.63-.67]
DV offenders	DV	.59	[.57-.62]	.59	[.57-.62]
Community					
All offenders	General	.75	[.74-.75]	.72	[.71-.72]
DV offenders	General	.73	[.71-.74]	.69	[.67-.70]
DV offenders	DV	.69	[.67-.71]	.67	[.65-.69]

Note: 95% CI = confidence interval.

For offenders in the community, the Community TRAS and LSI-R showed similarly acceptable predictive validity for general recidivism among the full sample. Discrimination performance declined slightly when assessing general recidivism among DV offenders, and again when assessing DV-related recidivism for DV offenders. Incremental losses in predictive validity when assessing DV recidivism were substantially smaller for the community sample, relative to the custody sample. AUC values for prediction of DV recidivism among DV offenders were consequently fair at .67 for the LSI-R and .69 for the Community TRAS.

## Understanding differences in predictive validity for DV recidivism

The pattern of AUC statistics indicated that risk assessments were particularly poor at predicting DV-related recidivism for DV offenders in custody. This was both relative to assessments' performance in predicting general recidivism, and also in comparison to performance in predicting DV-related recidivism for DV offenders in the community.

Observed differences between custody and community samples may not be attributable to variation in the assessment tools used (Custody TRAS vs Community TRAS) because similar trends were observed for the LSI-R, which has standard applications for both samples.

Exploring variability in the performance of risk assessment tools could provide insights into factors that influence predictive validity for DV-related recidivism. To this end we identified and tested possible sources of difference across samples and measures that may contribute to the observed pattern of discrimination statistics.

## Definitions of recidivism

One possibility is that definitions of DV recidivism used in the community sample (DV convictions resulting in a new custodial or supervised community episode) were easier to discriminate than definitions used in the custody sample (DV convictions resulting in a new custodial episode). By extension, this implies that assessment tools were less accurate at predicting return to custody than return to a new community order among DV offenders.

To assess this, we split the definition of recidivism in the community sample into separate outcomes for convictions resulting in a new custodial episode, and convictions resulting in a new community episode. We then tested the predictive validity of the LSI-R and Community TRAS against each of these definitions, with reference to DV-related offending among DV offenders.

AUC statistics for return to custody indicated acceptable discrimination performance at .76 (95% CI = .74-.78) for the Community TRAS and .74 (95% CI = .72-.76) for the LSI-R. These discrimination figures were higher than those for return to a new community episode outcomes when assessed by the Community TRAS (AUC = .65; 95% CI = .63-.67) or the LSI-R (AUC = .62; 95% CI = .60-.64) for community-based DV offenders. We also noted that discrimination performance for DV-related return to custody outcomes was consistently higher for DV offenders in the community sample than for those in the custody sample (see Table 2).

## Signal detection performance

Another possibility is that risk assessment tools could have different discrimination characteristics depending on the outcome being measured (e.g. general vs DV recidivism) and the target population (e.g. community vs custody samples). Discrimination statistics such as AUC are derived from signal detection measures of sensitivity and specificity. Sensitivity refers to accurate detection of individuals who return a signal, which in this instance is recidivism; specificity refers to accurate rejection of individuals who do not return that signal. Differences in the overall predictive accuracy of measures can therefore be a function of changes in their sensitivity and / or specificity.

Comparison of sensitivity and specificity statistics requires selection of a binary threshold, or series of binary thresholds, that can be consistently applied across outcomes and samples. This was not feasible for the Custody TRAS and Community TRAS because they are different measures and generate scores that are not comparable. We therefore used the example of the LSI-R total score threshold that warrants classification into the medium category of risk (total score  $\geq 24$ ). Historically, offenders have been required to have medium or higher risk on the LSI-R to be eligible for a number of programs delivered by Corrective Services NSW, including EQUIPs.

Table 3 gives signal detection statistics for the LSI-R medium risk threshold when estimating likelihood of DV recidivism among DV offenders, and when estimating likelihood of general recidivism among all offenders in the sample. AUC statistics indicated that the LSI-R threshold showed fair discrimination for all outcomes with the exception of DV recidivism in the custody sample, which was poor and only marginally better than chance.

In the custody sample, the LSI-R threshold had high sensitivity for DV offenders who return to custody with DV recidivism. This indicates that most DV recidivists had medium or higher risk on the LSI-R. However, specificity was extremely low, indicating that a large proportion of DV offenders with medium or higher risk did not return to custody as a result of DV-related recidivism. By comparison, the LSI-R threshold showed slightly lower sensitivity for general recidivism, although this was associated with marked improvements in specificity.

The LSI-R threshold showed different discrimination characteristics in the community sample, marked by lower sensitivity and higher specificity, compared to the custody sample. This suggests that a greater proportion of offenders who scored below the LSI-R threshold were recidivists, and a lower proportion of offenders above the threshold were not recidivists, in the community sample. Differences in discrimination between general and violent recidivism were similar to those of the custody sample, whereby the LSI-R threshold was slightly less sensitive although showed substantially better specificity for general recidivism compared to DV recidivism.

**Table 3.** Discrimination characteristics of the LSI-R medium risk threshold for general recidivism among all offenders, and DV recidivism among DV offenders, in the custody and community samples.

Sample	Recidivism	AUC	[95% CI]	sensitivity	specificity
Custody					
All offenders	General	.64	[.63-.65]	.91	.37
DV offenders	DV	.55	[.53-.58]	.95	.16
Community					
All offenders	General	.66	[.65-.67]	.60	.72
DV offenders	DV	.62	[.61-.64]	.67	.58

Note: 95% CI = confidence interval.

## CONCLUSIONS

This study aimed to explore whether risk assessment tools for general recidivism, including the LSI-R as well as the TRAS tools developed for local case management purposes by Corrective Services NSW, had offence-specific predictive validity for DV recidivism.

We found that for custody-based offenders, the LSI-R and Custody TRAS had acceptable discrimination performance when assessing general recidivism among the total sample. Performance declined slightly when assessing general recidivism risk among DV offenders, which may be related to subsample differences as well as the high incidence of specific offence categories (in this case, DV) in overall recidivism outcomes. Predictive validity was markedly poorer when assessing DV recidivism risk among DV offenders, and some measures such as the LSI-R medium risk threshold performed only marginally better than chance.

By comparison, changes in predictive validity across measures were more modest for community-based offenders. Discrimination statistics remained in fair to acceptable ranges when assessing accuracy of the LSI-R and Community TRAS for general recidivism among all offenders; general recidivism among DV offenders; and DV recidivism among DV offenders. AUC statistics for DV recidivism were .67 when assessed by the LSI-R and .69 when assessed by the Community TRAS, which are within predictive validity ranges previously reported for tools that are specifically designed to estimate DV risk (e.g. Graham et al., 2019; Hilton et al., 2010; Messing & Thaller, 2013).

Patterns of discrimination statistics across measures and samples provided some indications as to why risk assessment tools for general recidivism may vary in predictive validity, and underperform in some cases, when estimating risk of DV reoffending. Using the example of the LSI-R medium risk threshold, we found that assessments had poorer specificity for DV recidivism than general recidivism. This was particularly pronounced in the case of assessments for custody-based offenders, where more than 8 out

of 10 over the medium risk threshold were ‘false alarms’ or did not return to Corrective Services NSW for DV-related recidivism.

Studies have previously found that DV recidivists tend to have higher assessed risk and needs for general reoffending compared to DV non-recidivists (e.g. Cavanaugh & Gelles, 2005). This suggests that specificity is likely to be a particular challenge for general risk prediction tools, because a high risk classification would capture DV recidivists as well as more versatile offenders who may have a history of DV (among other) offending although do not have particular propensities towards repeated DV (Bouffard & Zedaker, 2016). In this case, the poor predictive validity of custody-based assessments for DV recidivism may have been exacerbated by indications that offenders in the custody sample had higher risk on average than offenders in the community sample.

On the other hand, the results of this and other studies suggest that DV offenders with low assessed general risk of recidivism are less likely to reoffend, both in terms of DV and overall. An implication is that general risk assessment tools may have utility in screening out low risk DV offenders and triaging higher risk DV offenders to more focused assessment of offence-specific risk and needs.

While there were indications that general risk assessments may perform better at discriminating lower risk DV offenders who do not reoffend, this pattern was complicated by different outcome measures used for the custody sample and the community sample. Interestingly, follow up analyses showed that assessments of community-based DV offenders had the strongest predictive validity for convictions resulting in a new custodial sentence, relative to new community orders alone or an aggregate of both outcomes. Imposition of a new community order is likely to have a lower threshold in terms of reoffending severity or frequency than imposition of a new custodial order. This could potentially result in decreased discrimination sensitivity as a result of increased ‘misses’, or instances of recidivism among offenders assessed as low risk. The results also suggest that challenges in estimating offence-specific recidivism using general risk assessment tools may not be solely attributable to statistical effects of lower base rates (e.g. Olver et al., 2014), given that community-based DV offenders were less likely to receive new custodial orders (10.4%) than new community orders (19.3%) for DV-related recidivism.

In sum, the results of this study indicate that general risk assessment tools can have moderate predictive validity for offence-specific recidivism among DV offenders. Performance varied substantially as a function of offender population and definitions of recidivism, however, and there were indications that predictive validity was only marginally better than chance in some conditions. It was therefore not possible to conclude that general risk assessments, and binary decision making thresholds such as having medium or higher risk on the LSI-R, have consistent applicability as a primary method of estimating offenders’ risk of DV recidivism and priority for offence-specific interventions. Conversely, the findings suggest that such assessments could have potential utility as a means of screening out DV offenders who are at low risk of serious DV-related recidivism.

## REFERENCES

---

- Andrews, D. A., & Bonta, J. (1995). *Level of Service Inventory-Revised (LSI-R): An offender assessment system. User’s guide*. Toronto, ON: Multi-Health Systems.
- Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct (5th ed)*. New Providence: LexisNexis Group.
- Bouffard, A., & Zedaker, B. (2016). Are domestic violence offenders specialists? Answers from multiple analytic approaches. *Journal of Research in Crime and Delinquency*, 53, 788–813.
- Carrington, K., & Phillips, J. (2006). *Domestic violence in Australia – An overview of the issues*. Canberra: Parliament of Australia.



- Cavanaugh, M. M., & Gelles, R. J. (2005). The utility of male domestic violence offender typologies: New directions for research, policy, and practice. *Journal of Interpersonal Violence, 20*, 155–166.
- Cox, P. (2015). *Violence against women in Australia: Additional analysis of the Australian Bureau of Statistics' Personal Safety Survey, 2012*. ANROWS.
- Graham, L. M., Sahay, K. M., Rizo, C. F., Messing, J. T., & Macy, R. J. (2019). The validity and reliability of available intimate partner homicide and reassault risk assessment tools: A systematic review. *Trauma, Violence, & Abuse, 15*(24838018821952).
- Hanson, R.K., Helmus, L., & Bourgon, B. (2007). *The validity of risk assessments for intimate partner violence: A meta-analysis*. Ontario: Public Safety Canada.
- Hendricks, B., Werner, T., Shipway, L., & Turinetti, G. J. (2006). Recidivism among spousal abusers: Predictions and program evaluation. *Journal of Interpersonal Violence, 21*, 703–716.
- Hilton, N., Harris, G., Popham, S., & Lang, C. (2010). Risk Assessment Among Incarcerated Male Domestic Violence Offenders. *Criminal Justice and Behavior, 37*, 815–832.
- Hilton, N. Z., Harris, G. T., & Rice, M. E. (2010). *Risk assessment for domestically violent men: Tools for criminal justice, offender intervention, and victim services*. Washington: American Psychological Association.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression (2nd ed.)*. New York: John Wiley & Sons.
- Laing, L., & Bobic, N. (2002). *Literature review: Economic costs of domestic violence*. Sydney: Australian Domestic & Family Violence Clearing House.
- Messing, J. T., & Thaller, J. (2013). The average predictive validity of intimate partner violence risk assessment instruments. *Journal of Interpersonal Violence, 28*, 1537–1558.
- Olver, M.E., Stockdale, K.C., & Wormith, J.S. (2014). Thirty years of research on the Level of Service scale: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment, 26*, 156–176.
- Ragusa-Salerno, L. M., Ostermann, M., & Thomas, S. S. (2013). Does the LSI-R have utility for sex offenders? *Criminal Justice and Behavior, 40*, 952–969.
- 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.
- 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.
- Zhang, Y., & Howard, M. (2020). *Understanding the spectrum of domestic violence: Risk factors, treatment pathways and recidivism among offenders who commit intimate partner or non-intimate partner violence*. Sydney: Corrective Services NSW.



© Corrective Services NSW

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](mailto:research.enquiries@justice.nsw.gov.au)