ADVANCING RESEARCH IN CORRECTIONS

School of Criminology and Criminal Justice | University of Nebraska Omaha
6001 Dodge Street, Omaha, NE 68182 · 402.554.2610 · arcorrectionslab.org

Evaluating the Minnesota Level of Service/Case Management Inventory

Research Summary

The Level of Service/Case Management Inventory (LS/CMI) is one of the most widely used risk and needs assessment tools in corrections. Developed in Canada, the LS/CMI is designed to predict recidivism risk and guide case management decisions. Research consistently shows that assessment tools perform differently when applied in new jurisdictions due to differences in population characteristics, policy environments, and implementation practices. As a result, best practice requires periodic local validation to ensure tools remain accurate and aligned with operational needs.

Using data from 131,899 assessments conducted in Minnesota between 2012 and 2022, along with qualitative input from probation staff across 43 agencies, this study evaluated the LS/CMI's usability, reliability, predictive validity, and functionality.

Findings indicate that the LS/CMI achieves weak-to-moderate predictive performance for general and violent recidivism. Yet the tool demonstrates excellent inter-rater reliability. However, the evaluation identifies consistent overprediction of risk, bias across demographic groups, and substantial variation in how the tool is implemented across counties.

The findings suggest that the LS/CMI's performance reflects both the strengths of practitioner implementation and the limitations of a tool that is not optimized for Minnesota's probation population. The results point to clear opportunities to improve accuracy, equity, and practical utility through recalibration, standardization, and consideration of locally developed assessments.

Risk-need assessment tools have become central to evidence-based corrections practice in the United States. Since the adoption of the Risk-Need-Responsivity (RNR) model in the late 1980s, correctional agencies have increasingly relied on actuarial instruments to guide supervision intensity, identify treatment targets, and allocate programming resources. These tools offer a significant improvement over unstructured clinical judgment, which research has shown to be inaccurate and susceptible to bias (Andrews & Bonta, 2010).

Third and fourth generation tools, such as the Level of Service/Case Management Inventory (LS/CMI), extended earlier criminal history measures by incorporating dynamic need items covering domains such as education, employment, family, and antisocial attitudes. These additions are intended to serve dual purposes: improving predictive accuracy and informing case planning decisions (Singh et al., 2018).

Despite widespread adoption, the performance of risk assessment tools is highly context dependent. Instruments developed and validated in one jurisdiction frequently demonstrate reduced predictive accuracy when applied elsewhere, a phenomenon known as predictive shrinkage (Hamilton et al., 2025). Local differences in population characteristics, policy, and supervision practices can erode a tool's accuracy and equity over time. Therefore, best practice requires periodic local validation to ensure continued fit for the population being assessed (Desmarais et al., 2022).

Minnesota presents a particularly complex environment for risk assessment. The state operates three distinct probation systems—Community Corrections Act agencies, the Minnesota Department of Corrections, and County Probation Offices—across 87 counties. This decentralized structure creates wide variation in training, cut points, and documentation practices. The LS/CMI has been used statewide since the early 2000s, but prior research has raised concerns regarding predictive accuracy and potential bias (Duwe & Rocque, 2016; Duwe, 2024). In response, the Minnesota Department of Corrections contracted the Advancing Research in Corrections (ARC) Lab at the University of Nebraska–Omaha to conduct a comprehensive evaluation.

Current study

This evaluation examined the LS/CMI's usability, reliability, and predictive validity across Minnesota's probation population. The study addressed three primary objectives:

- Assess the LS/CMI's predictive accuracy for general and violent reconviction, including calibration and differential prediction across gender and race/ethnicity.
- Evaluate implementation consistency through focus groups examining training, scoring, and operational practices across county agencies.
- Test inter-rater reliability (IRR) to determine whether scoring is consistent across assessors.

Data and methods

Data included LS/CMI item-level responses, domain scores, total scores, and risk level categories for 131,899 assessments completed between January 1, 2012, and March 31, 2022. These records were linked to demographic information, case characteristics, and reconviction outcomes defined as any new conviction within three years of assessment.

Predictive validity was assessed using two primary metrics. Discrimination was measured via the Area Under the Curve (AUC), which ranges from 0.5 (no better than chance) to 1.0 (perfect prediction); values between 0.64 and 0.70 indicate moderate discrimination, while values above 0.71 are considered strong (Rice & Harris, 2005). Odds ratios (ORs) are used to assess the magnitude of differences in reoffending likelihood. ORs close to 1.00 indicate little-to-no meaningful difference in reoffending likelihood, while values that depart further from 1.00 reflect stronger effects. ORs greater than 1.00 indicate an increased likelihood of reoffending; values below 1.00 indicate a decrease. Additionally, ORs represent effect sizes where 1 to 1.4 are negligible, 1.5 to 2.4 are small, 2.5 to 4.2 are moderate and 4.3 or greater are considered large effects.

Calibration was examined to assess whether predicted recidivism probabilities align with observed rates. Formal bias testing employed logistic regression models following the Cleary method, using group indicators and score-by-group interaction terms to detect intercept and slope bias across sex and race/ethnicity subgroups.

The qualitative component consisted of five web-based focus groups with managers and line staff from 43 agencies statewide. Sessions were recorded, transcribed, and analyzed using thematic analysis procedures to capture implementation experiences, scoring challenges, and operational practices. Inter-rater reliability was evaluated using linear weighted kappa statistics from a dedicated proficiency testing dataset.

Key findings

1. Usability: Wide Variation Across Counties

Focus group participants described the LS/CMI as a valuable framework for communicating risk but emphasized that its application varies widely across Minnesota's counties. Differences in training, scoring interpretation, cut points, and documentation practices have produced notable inconsistencies across agencies.

Participants also highlighted challenges with ambiguous item wording when scoring the tool, particularly in the Companions and Leisure/Recreation domains. Additionally, the legalization of THC in Minnesota has introduced further variability, as agencies rely on local standards to interpret drug-related items. To address perceived limitations of the LS/CMI, practitioners often supplement it with trailer tools such as the DRI, IDA, and WRNA, especially for cases involving DUI, domestic violence, sex offenses, and female populations. These inconsistencies are especially problematic during intrastate transfers, where differences in instruments and risk thresholds across counties undermine the portability and comparability of risk scores.

2. Inter-Rater Reliability: Excellent Agreement

Assessors demonstrated a high degree of consistency overall, with results indicating excellent reliability. At the item level, agreement was nearly perfect across all 43 items. Criminal history items showed the strongest agreement, reflecting their basis in objective records, but even needs-based items were scored with similarly strong consistency, despite concerns raised in the focus groups. Together, the findings suggest that performance challenges are less driven by scoring inconsistency and instead point to potential limitations in how the tool functions within Minnesota.

3. Predictive Validity: Moderate to Weak

Table 1 displays the predictive validity results for general and violent recidivism across demographic groups. Overall, the tool demonstrated moderate predictive validity for general reconviction (AUC = 0.65) and weak performance for violent reconviction (AUC = 0.63). Predictive accuracy was largely consistent across gender, with modest variation by race and ethnicity; the lowest accuracy was observed among American Indians, with AUC values of 0.60 for general and 0.59 for violent reconviction.

Table 1. AUC Values for General & Violent Reconviction

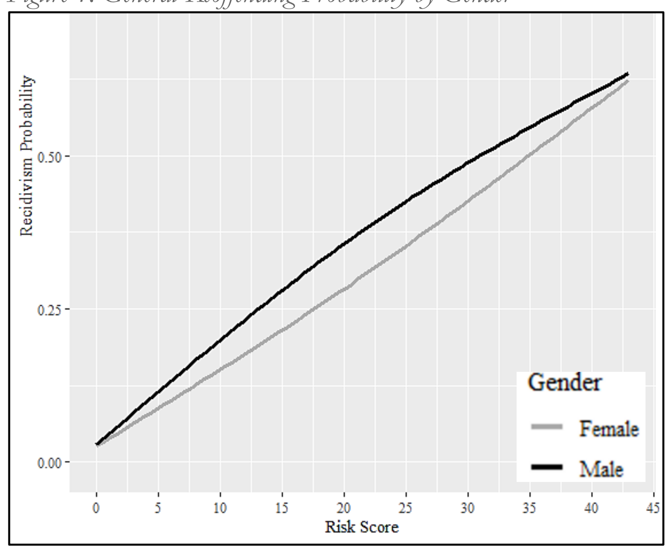
Group	General	Violent
Total	0.65	0.63
Male	0.65	0.65
Female	0.64	0.63
White	0.66	0.63
Black	0.62	0.62
Hispanic	0.66	0.65
American Indian	0.60	0.59
Other/Unknown	0.65	0.63

Criminal history and needs domains produced nearly identical predictive value (AUC = 0.63 and 0.64, respectively), suggesting that the needs components add little incremental value beyond criminal history alone. The Criminal History and Antisocial Pattern domains were the strongest predictors of reoffending, while most other areas added little additional accuracy.
Moreover, results confirmed concerns raised in focus groups, where post-transfer assessments performed notably worse (AUC = 0.62) than initial assessments (AUC = 0.66), highlighting the impact of cross-county inconsistencies on predictive performance.

4. Prediction Bias: Modest but Consistent

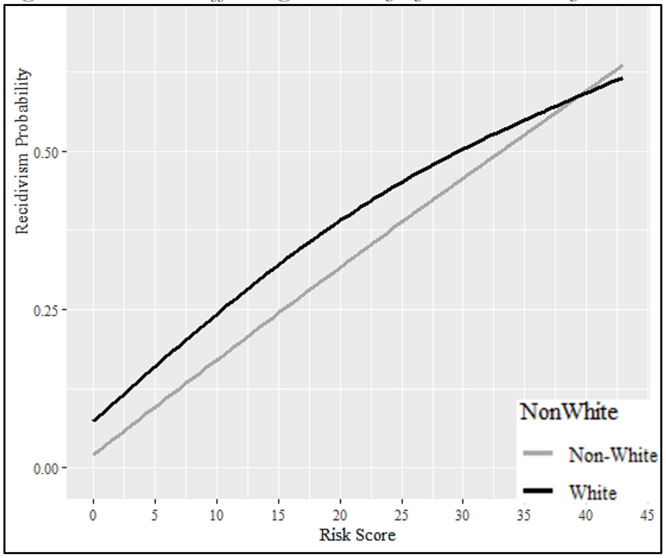
Formal bias testing found that the tool overestimates risk for certain groups. Figure 1 shows that women with the same LS/CMI scores as men were less likely to reoffend, meaning they are more often classified as higher risk than their score would predict. This gap becomes even more pronounced as scores increase, where men were up to 11% more likely to reoffend than women at the same level.

Figure 1. General Reoffending Probability by Gender



A similar pattern appears across race/ethnicity, shown in Figure 2. Non-White individuals generally reoffended at lower rates than White individuals with the same LS/CMI scores, indicating that the tool overestimates risk for these groups. Items in the Education/Employment and Companions domains were especially more likely to flag non-White individuals as higher risk even when they did not go on to reoffend, contributing to these disparities.

Figure 2. General Reoffending Probability by Race/Ethnicity

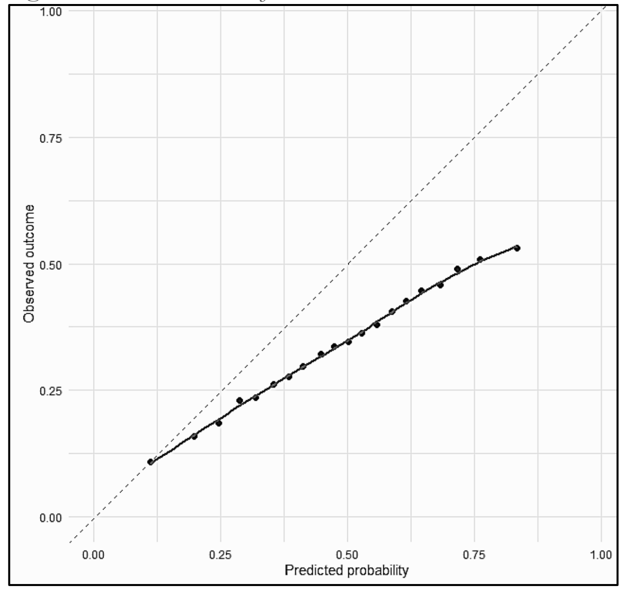


5. Functionality: Overprediction and Classification

Calibration analyses showed that LS/CMI scores overestimate reoffending patterns. This is shown in Figure 3, where the diagonal dotted line indicates that the predicted likelihood of reoffending perfectly matches observed reoffending rates. As the calibration trend drifts below the dotted line, this finding indicates that the tool's predicted probability does not match individuals' observed rate of recidivism. Specifically, those with higher risk scores are predicted to recidivate at higher rate than observed – overestimation. For example, 75% of people with a predicted probability of 0.75 (which equates to scoring 32 on the LS/CMI) should reoffend; however, Figure 3 shows that only 51% do – a 24% difference.

At the county level, risk level classification schemes show a clear pattern where higher risk categories are associated with higher rates of reconviction (with average ORs ranging from 2.1 for Low/Moderate to 10.2 for High/Very High when compared to Very Low/Low), suggesting the tool distinguishes between levels of risk. However, the size of these differences varies among agencies, and standardizing cut points to the classification scheme used by the DOC across counties would slightly reduce the contrasts between risk levels, creating less meaningful distinctions in reoffending probability between risk-level categories.

Figure 3. Calibration Plot for Minnesota LS/CMI Scores



Discussion

The evaluation reveals a system that has achieved notable success in implementing a risk assessment tool across a decentralized environment. The strong inter-rater reliability results reflect meaningful institutional investment in training and quality assurance. Focus group participants demonstrated nuanced understanding of the tool's capabilities and limitations, often identifying areas of concern that were subsequently confirmed through formal analyses.

At the same time, three consistent limitations emerged. First, the LS/CMI's modest predictive performance and bias reflects the challenges of applying tools developed elsewhere to a new population. Second, calibration analyses confirm that standard LS/CMI probabilities overestimate recidivism risk for Minnesota probationers, contributing to overclassification. Third, the decentralized structure of Minnesota's probation system amplifies the impact of tool limitations, as inconsistent cut points and trailer tool practices reduce the portability of risk scores across county lines.

Recommendations

The evaluation identifies clear pathways for improvement:

1. Adopt Recalibrated Cut Points

The recalibrated LS/CMI thresholds and probability estimates developed in this study should be adopted statewide to reduce overclassification and better align risk categories with Minnesota's observed recidivism rates. Recalibration requires no changes to item scoring and can be implemented without replacing the current system. This change would reduce supervision burdens while maintaining the tool's predictive structure.

2. Pursue Targeted Modifications to the LS/CMI

Where feasible and within proprietary constraints, item weights should be adjusted to improve prediction and reduce bias. Items that demonstrate weak predictive value or high false positive rates for specific demographic groups should be evaluated for removal or reweighted. Separate models for general and violent outcomes should be considered, as the current tool performs meaningfully worse for violent reconviction.

3. Evaluate Alternative Assessment Tools

The state should consider evaluating alternative risk assessment tools that offer greater flexibility and built-in gender-responsive and violence-specific models, but any option would need to be carefully calibrated to Minnesota's population. Even widely used proprietary tools like COMPAS or ORAS would face similar limitations if adopted without local validation. Newer tools such as MnSTARR or STRONG-R, which allow for adjustable scoring and population-specific models, may be better suited to the state's needs.

Alternatively, Minnesota is well positioned to develop its own assessment tool. With strong administrative data, demonstrated scoring reliability, and institutional experience, a locally designed instrument could be tailored to the state's probation population, incorporate gender-responsive and violence-specific components, and reduce reliance on supplemental tools. While this approach would require sustained investment in development and validation, it offers the *greatest potential for long-term improvements* in accuracy, fairness, and cost efficiency.

4. Strengthen Training and Quality Assurance

Regular inter-rater reliability checks, refresher training, and verification that all agencies are using current scoring manuals are essential to maintain the scoring consistency observed in this study. A single, unified Minnesota-specific scoring handbook incorporating current guidance and state-specific contexts would reduce confusion in scoring specific items.

5. Standardize Implementation While Acknowledging Resource Constraints

Clear statewide guidance for scoring, documentation, and risk category thresholds should be established, with recognition that agencies have different resource environments. Standardization should be accompanied by resources to address gaps between supervision standards and actual caseload capacity. Consistent S^3 system documentation should be made mandatory to support portability of risk information across counties.

Conclusion

Risk assessment tools like the LS/CMI offer significant advantages over unstructured decision-making, but only when they perform accurately and equitably for the populations they serve. Minnesota's evaluation demonstrates both the capacity of the state's probation system to implement risk assessments at scale and the limitations that emerge when tools developed elsewhere are applied in a new context without local calibration.

The findings support continued use of the LS/CMI in the near term, with recalibrated thresholds and continued training and quality assurance. In parallel, the state should evaluate alternative tools and consider local development as a longer-term strategy to improve prediction, reduce overclassification, and better serve Minnesota's diverse probation population. With the infrastructure, data, and institutional knowledge already in place, Minnesota is well positioned to build an evidence-based risk assessment that advances both accuracy and equity in community supervision.

References

Andrews, D. A., & Bonta, J. (2010). The psychology of criminal conduct. Routledge.

Council of State Governments Justice Center. (2023). Justice reinvestment initiative in Minnesota: Improving supervision investments and outcomes. https://csgjusticecenter.org

Desmarais, S. L., D'Amora, D. A., & Tavárez, L. P. (2022). Advancing fairness and transparency: National guidelines for post-conviction risks and needs assessment. U.S. Department of Justice, Office of Justice Programs.

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.

Duwe, G. (2024). Evaluating bias, shrinkage and the home-field advantage: Results from a revalidation of the MnSTARR 2.0. Corrections, 9(1), 20–42.

Hamilton, Z., Kigerl, A., Allen, B., Ursino, J., & Krushas, A. (2025). Never going to let you down: Preventing predictive shrinkage via the STRONG-R assessment method. Justice Quarterly, 42(7), 1279–1299.

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(1), 156.

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(5), 615–620.

Singh, J. P., Kroner, D. G., Wormith, J. S., Desmarais, S. L., & Hamilton, Z. (Eds.). (2018). Handbook of recidivism risk/needs assessment tools. John Wiley & Sons.

Skeem, J. L., & Lowenkamp, C. T. (2016). Risk, race, and recidivism: Predictive bias and disparate impact. Criminology, 54(4), 680–712.

Wildermuth, J., Nonemaker, D., & Carlson, J. (2021). Validation of the Level of Service Case Management Inventory (LSCMI) for assessing risk of re offense in Hennepin County. Hennepin County Department of Community Corrections and Rehabilitation.

Suggested citation: Hamilton, Z., Ursino, J., & Kim, Y. (2025). Validating the Minnesota Level of Service/Case Management Inventory (LS/CMI) – Research Brief (Doc. #2025-003). Advancing Research in Corrections Lab. Omaha, NE.