

A Systematic Review of the Predictive Validity of the VRAG-R

Vivian Au, MD, Aariz Naeem, Paul Benassi, MD, MSc, Sarah Bonato, and Roland M. Jones, PhD, MSc, MB ChB, BSc

Structured risk assessment tools are essential in forensic psychiatry to evaluate the likelihood of recidivism. The Violence Risk Appraisal Guide-Revised (VRAG-R) was developed as an update to the VRAG, but its predictive validity across offender populations remains underexamined. Our study aimed to examine the predictive validity of the VRAG-R for general, violent (including and excluding sexual offenses), and sexual recidivism. We conducted a systematic review and meta-analysis, searching 10 databases and gray literature sources for studies reporting psychometric outcomes for the VRAG-R published since 2013. Risk of bias was assessed using Prediction Model Risk of Bias Assessment (PROBAST) and data extraction followed the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies (CHARMS) checklist. Area under the curve (AUC) values were pooled using random-effects meta-analysis. In total, 15 studies comprising 3,932 participants were included. The VRAG-R showed acceptable predictive validity for general recidivism (pooled AUC = .71, 95% CI: .67 to .75) and violent recidivism (AUC = .72, 95% CI: .69 to .75). Predictive validity for sexual recidivism was modest (AUC = .65, 95% CI: .61 to .68). In conclusion, the VRAG-R demonstrates acceptable predictive validity for general and violent recidivism, comparable with other tools. Its performance in predicting sexual recidivism, however, is limited and concerns about generalizability remain. Future research should prioritize diverse samples, reporting of calibration, and continued evaluation of performance.

J Am Acad Psychiatry Law 54(1) online, 2026. DOI:10.29158/JAAPL.260001-26

Key words: actuarial risk assessment; aggression and violence; predictive validity; recidivism; risk assessment; Violence Risk Appraisal Guide-Revised; VRAG-R

Assessing the risk of criminal recidivism is a crucial task in forensic psychiatry that directly informs the practices of the criminal justice system to protect the public. Over time, the methodological approach for risk assessment has shifted toward structured professional judgment (SPJ) and actuarial tools, moving away from unstructured clinical judgment, which relied solely on individual clinical impressions.^{1,2} Actuarial risk assessment relies on both discrimination, to distinguish between high- and low-risk individuals, and

calibration, to ensure the predicted probabilities reflect the observed outcomes. In forensic risk assessment, the area under the receiver operating characteristic curve (AUC) is used to assess a tool's ability to distinguish between individuals who do or do not recidivate, with values ranging from .5 to 1. For example, an AUC of .5 indicates that the tool is no better than chance, whereas an AUC of 1 indicates that the tool perfectly distinguishes individuals each time.³ The authors of the VRAG tool have characterized an AUC of .56 as small, .64 as moderate, and .71 or higher as high.⁴ In contrast, SPJ involves a combination of actuarial methods and unstructured clinical judgment, allowing clinicians to consider defined risk factors while also exercising professional judgment in weighing and integrating them into the overall risk determination.

The Violence Risk Appraisal Guide (VRAG) was a pioneering tool in the field, as it was among the first actuarial instruments developed. Clinicians completed a 12-item scale that placed an offender in one of nine risk categories. The scale was developed based

Published online March 2, 2026.

Dr. Au is a PGY5, Dr. Benassi is an assistant professor, and Dr. Jones is an associate professor, Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada. Mr. Naeem is a Bachelor of Science candidate, Faculty of Life Sciences, McMaster University, Hamilton, Ontario, Canada. Dr. Benassi and Dr. Jones are psychiatrists, Forensic Mental Health Division and Dr. Jones is a scientist, Institute for Mental Health Policy. Ms. Bonato is a librarian, Centre for Addiction and Mental Health, Toronto, Ontario, Canada. Address correspondence to: Vivian Au, MD. E-mail: vivian.au@medportal.ca.

Disclosures of financial or other potential conflicts of interest: None.

on offender characteristics that were found to most strongly correlate with violent recidivism.⁵ Although the instrument was originally developed using a sample of male offenders who were assessed or treated at a maximum-security psychiatric facility in Ontario, Canada, its findings have since been replicated across a range of populations internationally. Additionally, the tool has been extensively validated, with more than 60 successful replication studies since its creation demonstrating an average AUC of .72.^{6,7}

Since the original release of the VRAG, the field has seen a proliferation in the development of risk assessment tools.^{8,9} This growth reflects the emerging recognition of the limited accuracy of unstructured clinical judgment, which has been shown to be unreliable and prone to biases.² Moreover, the superiority of structured and standardized prediction methods has been consistently replicated in forensic research.¹ Today, more than 200 structured tools are utilized within criminal justice systems globally, providing clinicians with an abundance of options but also creating challenges in determining which tool is most appropriate to use.^{10,11}

Despite these developments, the VRAG has maintained its status as one of the most commonly used risk assessment tools in the field.¹⁰ In 2013, the Violence Risk Appraisal Guide-Revised (VRAG-R) was released, incorporating a larger sample that included most of the original cohort. The purpose of the VRAG-R was to simplify the scoring system of the VRAG by replacing the item requiring the total Psychopathy Checklist-Revised (PCL-R)¹² score with only Facet 4 of the PCL-R.⁶ Additionally, the authors introduced an item that evaluated sexual offending, recommending the VRAG-R as a replacement for both the VRAG and the Sex Offender Risk Appraisal Guide (SORAG),^{7,13} which had been developed as a modification of the VRAG to predict recidivism among sexual offenders.¹⁴

It is important to examine the predictive validity of the VRAG-R in its own right, as the adoption has likely been influenced by the extensive validation of its predecessor.⁶ This systematic review and meta-analysis aims to critically evaluate the predictive performance of the VRAG-R. We specifically assess whether the VRAG-R demonstrates predictive accuracy comparable with that of the original VRAG across diverse offender populations and for different types of recidivism, including violent, sexual, and general reoffending. Additionally, where reported in studies of the VRAG-R, we compare the VRAG-R's performance against some of the most validated and

commonly used risk assessment tools, including structured professional judgment tools, such as the Historical Clinical Risk Management-20, Version 3 (HCR-20V3) and Sexual Violence Risk-20 (SVR-20), as well as actuarial tools, such as the Psychopathy Checklist-Screening Version (PCL:SV) and STATIC-99R.^{15,16} By synthesizing the existing literature, this study seeks to critically examine the tool's performance, which can assist practitioners and policymakers in better understanding the strengths and limitations of the VRAG-R in forensic risk assessment and how the tool performs compared with other available tools.

Methods

We developed a protocol following the Preferred Reporting Items for Systematic Review and Meta-Analysis guideline (PRISMA),¹⁷ which was registered on the International Prospective Register of Systematic Reviews PROSPERO (CRD42024599060). Our study used the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies (CHARMS)¹⁸ to guide data extraction and critical appraisal of our included studies.¹⁸ Furthermore, we used the Prediction Model Risk of Bias Assessment (PROBAST)¹⁹ to review each included study for risk of bias and applicability.¹⁹

Search Strategy

A systematic search strategy was designed in conjunction with a librarian experienced in systematic review searching (SB). We searched Medline, Criminal Justice Abstracts, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Embase, Education Resources Information Center (ERIC), Health and Psychosocial Instruments (HaPI), Mental Measurements Yearbook, PsycINFO, Scopus, and Web of Science for articles published from 2013 onward. No study type or language limits were applied to the search results. In addition, we carried out a manual review of the citations of included studies, a search of Google Scholar limited to the first 100 references, and a gray literature search of conference papers, dissertations, and preprints.

Study Selection Criteria

All studies that reported psychometric data on the VRAG-R published in English were eligible for inclusion, including cohort, case-control, and observational studies. Studies of populations of people convicted of a criminal offense were included, with

no exclusions based on gender, age, setting, or type of offense.

All identified references were uploaded to Covidence²⁰ to facilitate reference management. Two of the study authors (VA and AN) independently screened all titles and abstracts. Full-text review was completed by both reviewers only if one or both reviewers indicated the study met eligibility criteria based on preliminary review of the title and abstract. Any discrepancies at each stage were resolved through discussion and consensus between the two reviewers, consulting a third reviewer where necessary.

Data Extraction and Coding

Two of the study authors (VA and AN) independently extracted data using the CHARMS Checklist template.^{18,19} For all eligible articles, study information and design, participant characteristics, outcome measures (including general, violent, or sexual recidivism), VRAG-R performance (i.e., calibration, discrimination, and overall measures), and interpretation of the predictive validity of the VRAG-R were extracted. Within the calibration and discrimination measures, we additionally aimed to extract the observed: expected ratio (O:E) and the hazard ratio (HR) where available. The O:E refers to the number of recidivists actually observed to the number of recidivists predicted by the VRAG-R, whereas the HR compares the risk of recidivism between two groups.²¹

Risk of Bias Assessment

Two of the study authors independently completed an assessment of the methodological quality of the studies according to the four domains of PROBAST using the template created by Fernandez-Felix *et al.*¹⁹ The studies were examined for the risk of bias introduced by the inclusion and exclusion of participants as well as the definition and assessment of predictors and outcomes. Additionally, the articles were assessed for their methods of analysis, including their handling of missing data, reporting of results, and statistical methods.

Statistical Methods Used for Meta-Analysis

For a comprehensive analysis of the VRAG-R's performance, we aimed to examine both its discrimination and calibration measures. To enable consistent scaling across studies, we performed a logit transformation of the AUC values. Where data were available, we carried out separate meta-analyses for

different follow-up periods to explore temporal effects on VRAG-R performance.

Meta-Analysis

To quantitatively synthesize the predictive validity of the VRAG-R, we conducted meta-analyses on reported area under the curve (AUC) values across eligible studies. AUC was chosen as the primary performance metric because it was the only metric consistently reported across included studies. Although we intended to summarize calibration metrics, these were infrequently or inconsistently reported, precluding meaningful meta-analytic synthesis.

All meta-analyses were conducted using Stata (version 17.0; StataCorp LLC). AUC values were logit-transformed prior to analysis to stabilize variances and normalize distributions. Pooled estimates were calculated using a random-effects model (DerSimonian and Laird method)²² to account for expected heterogeneity across studies. Between-study heterogeneity was assessed using the I^2 statistic, with values above 50 percent interpreted as moderate to high heterogeneity.

To examine the influence of follow-up duration, we conducted subgroup analyses stratifying studies by follow-up length (less than five years versus greater than or equal to five years). Because of limited sample sizes, formal subgroup analyses by gender, age, or setting were not feasible.

Publication bias was visually assessed for asymmetry using funnel plots and statistically tested with Egger's test²³ where applicable.

Results

Our literature search identified 1,050 records, of which 429 duplicates were identified and removed by Covidence. A total of 621 titles and abstracts were screened, which excluded 569 studies. After reviewing the full text of 52 studies, 15 studies met inclusion criteria (Fig. 1) for the systematic review. Of these, 13 were retrospective cohort studies and two were prospective cohort studies. In total, 3,932 participants were included in the analysis, with most participants being adult White males. The studies were conducted across seven countries, including Canada, Switzerland, Austria, Mexico, Germany, Belgium, and Australia, which all have different criminal justice systems and population compositions. Only one study focused exclusively on female participants

A Systematic Review of the Predictive Validity of the VRAG-R

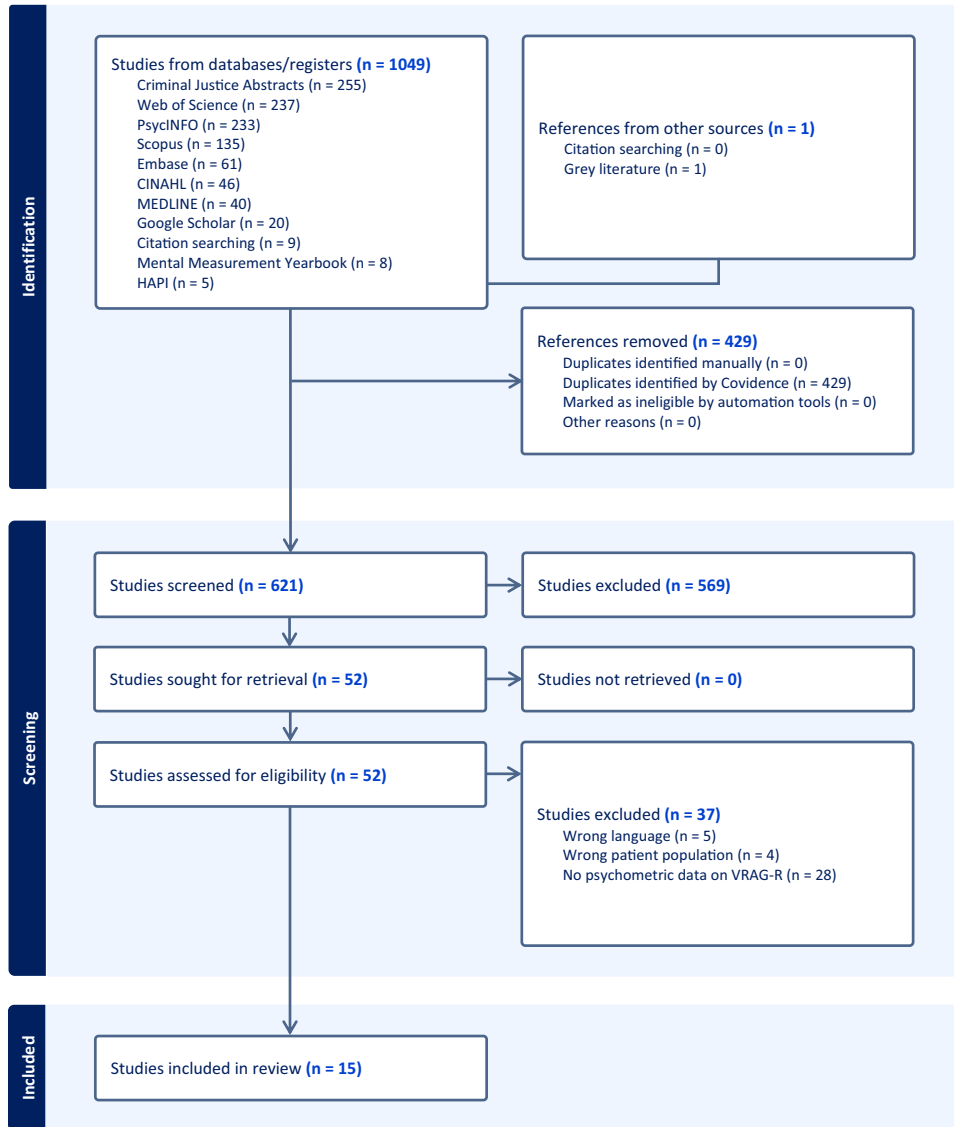


Figure 1. PRISMA flow diagram.

($n = 525$) and three on youth samples below the age of 25 ($n = 822$).

Recidivism was most often defined as any new criminal charges or conviction following a previous conviction. Recidivism data were extracted from various sources depending on the scope of the study, including federal and provincial databases or at the intuitional level, such as correctional facilities, hospitals, or programs tailored for specific populations (i.e., sex offenders and individuals found to be not criminally responsible.) There were three studies that focused solely on participants who were in institutional settings. For these participants, recidivism was defined as violent incidents as per the Modified Overt

Aggression Scale (MOAS) or the Staff Observation Aggression Scale-Revised (SOAS-R), although only incidents that were threatening or did harm to others were coded. For participants who were in the community, recidivism definitions differed across studies, particularly violent recidivism, where there was no consistency between studies as to whether sexual offenses were included in the definition. For example, although some studies categorized contact sexual offenses as violent recidivism,^{24–29} others explicitly excluded them from this category.^{1,30–35} To account for this variability, recidivism outcomes were categorized to include general recidivism, nonsexual violent recidivism, violent and sexual recidivism, as well as

sexual recidivism. There were three studies conducted with solely sexual offenders,^{26,29,31} examining the predictive validity of the VRAG-R for general, violent, and sexual recidivism.

All 15 studies reported the AUC as a measure of the ability of the tool to distinguish between recidivists and nonrecidivists. The follow-up times in all 15 studies varied, ranging from 28 days to 17.75 years. We carried out a meta-analysis of reported AUCs in these studies.

Across the 15 included studies, the AUCs of the VRAG-R varied depending on the recidivism category and follow-up duration. Focusing on long-term predictive validity, the pooled AUC values using the longest available follow-up period reported in each study were examined, according to the categories of general, nonsexual violent, violent and sexual, and sexual recidivism (Fig. 2).

For general recidivism, 10 studies were included, with AUC values ranging from .60 to .86 and a pooled estimate of .71 (95% confidence interval (CI) .67 to .75; $I^2 = 68.6\%$). For nonsexual violent recidivism, eight studies were included, with AUC values ranging from .60 to .80 and a pooled estimate of .72 (95% CI .68 to .76; $I^2 = .0\%$). Similarly, seven studies were included for violent and sexual recidivism, with AUC values ranging from .66 to .75 and a pooled estimate of .72 (95% CI .69 to .75; $I^2 = .0\%$). Finally, for sexual recidivism, five studies were included, with AUC values ranging from .63 to .69 and a pooled estimate of .65 (95% CI .61 to .68; $I^2 = .0\%$).

The degree of heterogeneity varied across outcome categories. For general recidivism, moderate heterogeneity was observed ($I^2 = 68.6\%$), suggesting substantial variability in effect sizes across studies. This may be attributed to differences in study populations, definitions of recidivism, or follow-up durations. In contrast, heterogeneity was negligible for nonsexual violent recidivism ($I^2 = .0\%$), violent and sexual recidivism ($I^2 = .0\%$), and sexual recidivism ($I^2 = .0\%$), indicating consistent effect sizes across studies for these outcomes.

To explore potential sources of heterogeneity, particularly for general and sexual recidivism, we conducted subgroup analyses based on follow-up duration (less than five years versus greater than or equal to five years). In our secondary analysis, the pooled AUC values for general, violent and sexual, and nonsexual violent recidivism were stratified by follow-up duration (Figs. 3 and 4). These analyses suggested that follow-up time did not account for the heterogeneity observed in

general recidivism, and pooled AUC values remained consistent across time points.

Although there were insufficient data to support formal subgroup analyses for specific populations, some insights can be taken from the available data. In the single study that focused on female forensic psychiatry patients, the VRAG-R was found to have an AUC of .69 (95% CI .65 to .74) for general recidivism and .68 (95% CI .61 to .75) for violent recidivism.⁴ Particularly, in Canadian studies, there was also minority representation of women in the sample, with one study concluding that women had lower VRAG-R scores in general when compared with the men.²⁸ There were three studies that solely focused on youth offenders, although the age definition of youth varied. Two of the three studies found that the VRAG-R had acceptable predictive validity for violent and general recidivism in youth whereas the other did not, with an AUC value of only .60 (95% CI .47 to .73) for general recidivism and an AUC value of only .63 (95% CI .50 to .77) for violent recidivism.^{27,30,34} Of the three studies focused on youth offenders, only one study examined sexual recidivism in youth sexual offenders, which reported an AUC of .69 (95% CI .53 to .85).³⁰ There were two studies that examined the predictive validity of the VRAG-R for inpatient violence, which is a population on which the VRAG was previously validated. In contrast to the VRAG, neither study found that the VRAG-R demonstrated predictive validity for inpatient violence.^{32,33} Finally, the predictive validity of the VRAG-R for sexual recidivism in sexual offenders was examined through three studies. All three concluded that the VRAG-R had poor to fair predictive validity for sexual recidivism, with AUC values ranging from .56 to .63.^{26,29,31}

All studies were rated as having high risk of bias overall using PROBAST, primarily because of deficiencies in the analysis domain. There was incomplete reporting of calibration measures, with only two studies reporting a Hosmer-Lemeshow test.³⁶ Additionally, a funnel plot was generated to assess for potential publication bias across the included studies (Fig. 5). The mild asymmetry, particularly among smaller studies, may indicate that studies with weaker predictive validity of the VRAG-R were not published or included.

Discussion

This systematic review and meta-analysis examined the predictive validity of the VRAG-R for general, violent, and sexual recidivism. The results indicated

A Systematic Review of the Predictive Validity of the VRAG-R

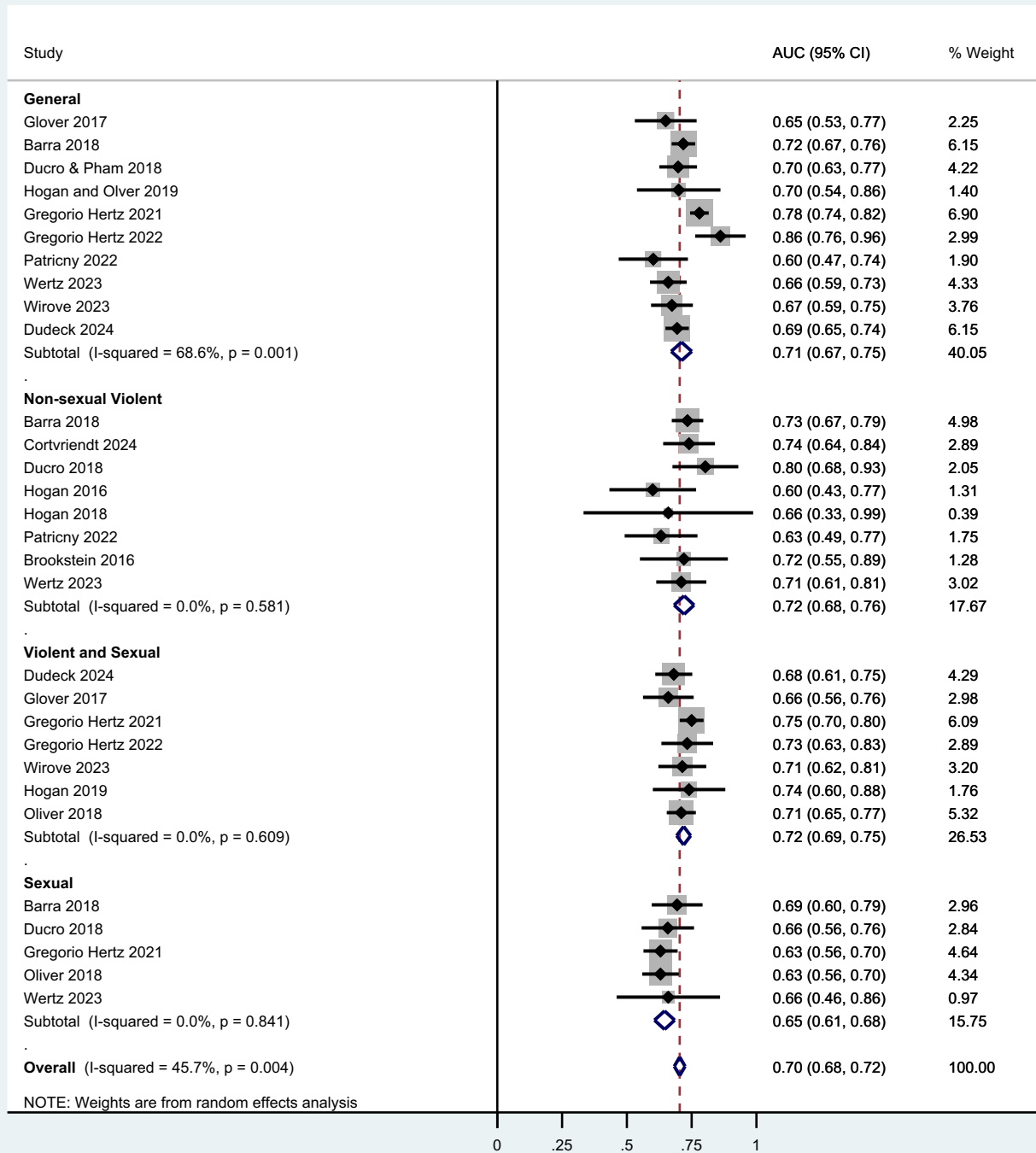


Figure 2. Meta-analysis of the predictive validity of the VRAG-R for general, nonsexual violent, violent and sexual, and sexual recidivism for the longest availability follow-up periods. AUC = area under the curve; CI = confidence interval.

acceptable to good predictive validity of the VRAG-R for general and violent recidivism, with pooled estimated AUC values falling between .67 and .75 for general recidivism and between .68 and .76 for violent recidivism. Contrary to the developmental sample, the

VRAG-R demonstrated only poor to fair predictive validity for sexual recidivism, with a pooled estimated AUC between .61 and .68.

To contextualize our findings, it is important to compare the pooled estimated AUC values for recidivism

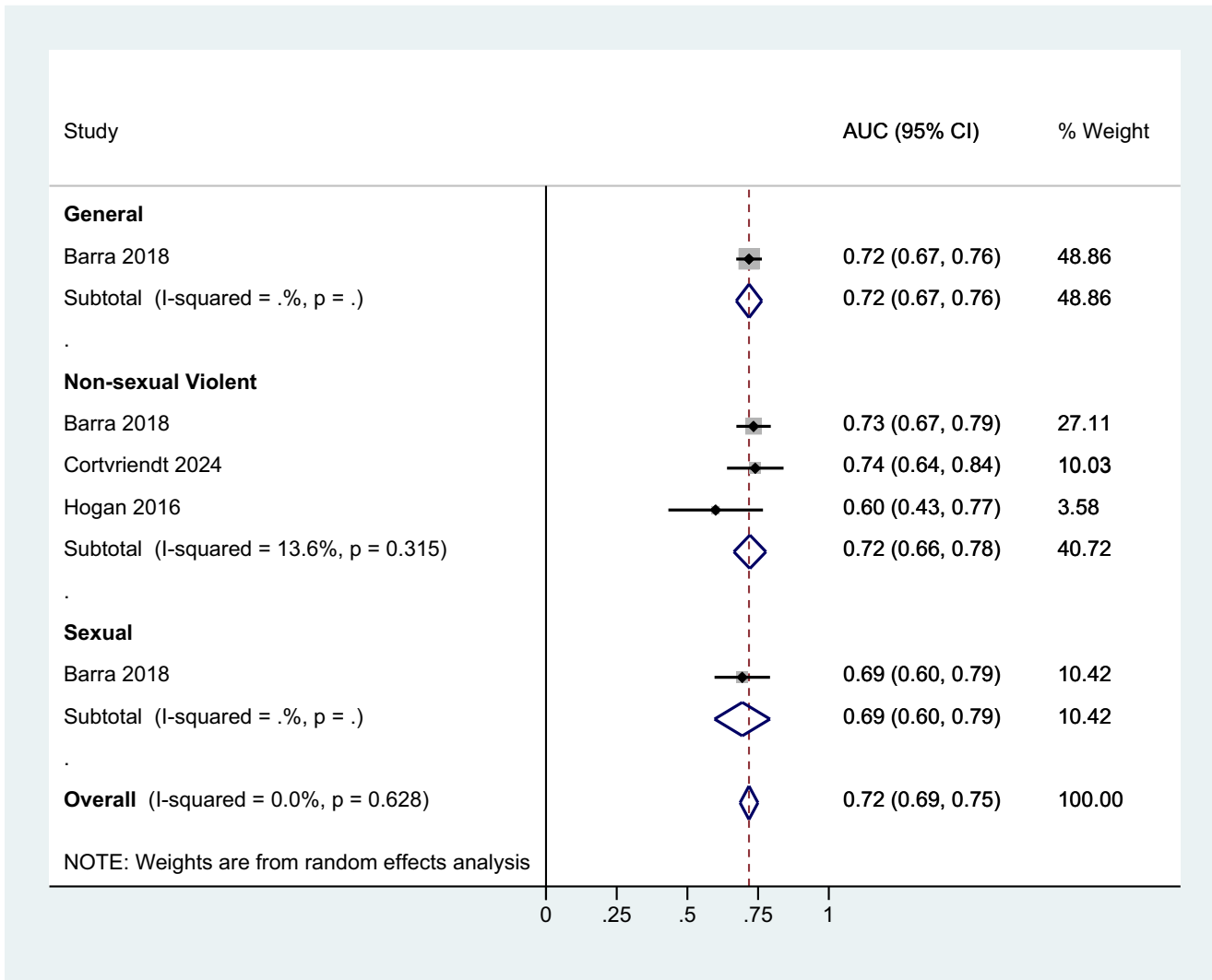


Figure 3. Meta-analysis of the predictive validity of the VRAG-R for general, nonsexual violent, violent and sexual, and sexual recidivism for less than five years of follow-up. AUC = area under the curve; CI = confidence interval.

with other commonly used risk assessment instruments. For general recidivism, our pooled estimated AUC across studies based on the longest follow-up periods available was .71 (95% CI .67 to .75; $I^2 = 68.6\%$). This estimate is similar to the predictive validity of other commonly used actuarial risk assessment tools, such as the HCR-20V3 and the PCL-SV, which have pooled AUC values of .69 (95% CI .65 to .72) and .67 (95% CI .56 to .77).¹⁶ This finding is replicated when examining violent (including sexual offenses) recidivism, where our pooled estimated AUC across studies based on the longest follow-up periods available was .72 (95% CI .69 to .75; $I^2 = .0\%$), which is similar to the pooled estimated AUC values for the HCR-20V3 (.69, 95% CI .65 to .72), Static-99 (.64, 95% CI .53 to .73), and the original VRAG (.69, 95% CI .63 to .75).¹⁶

The pooled estimated AUC based on the longest follow-up periods available for sexual recidivism was more modest, at .65 (95% CI .61 to .68; $I^2 = .0\%$). This estimate is in keeping with the established predictive validity of other commonly used instruments for predicting sexual recidivism. The Static-99 has a reported AUC value of .66 (95% CI .57 to .74), the SORAG ranges from .64 to .66, and the STABLE-2007 has an AUC of .67 (95% CI: .65 to .70).^{16,37-40} The SVR-20 has the highest predictive validity, with reported AUC values ranging from .72 to .80.⁴¹⁻⁴³ This comparison highlights a broader pattern among actuarial risk assessment tools for sexual recidivism, including the VRAG-R, that their predictive accuracy is generally modest.

Overall, the clinical implications of our findings suggest that the VRAG-R performs similarly to other

A Systematic Review of the Predictive Validity of the VRAG-R

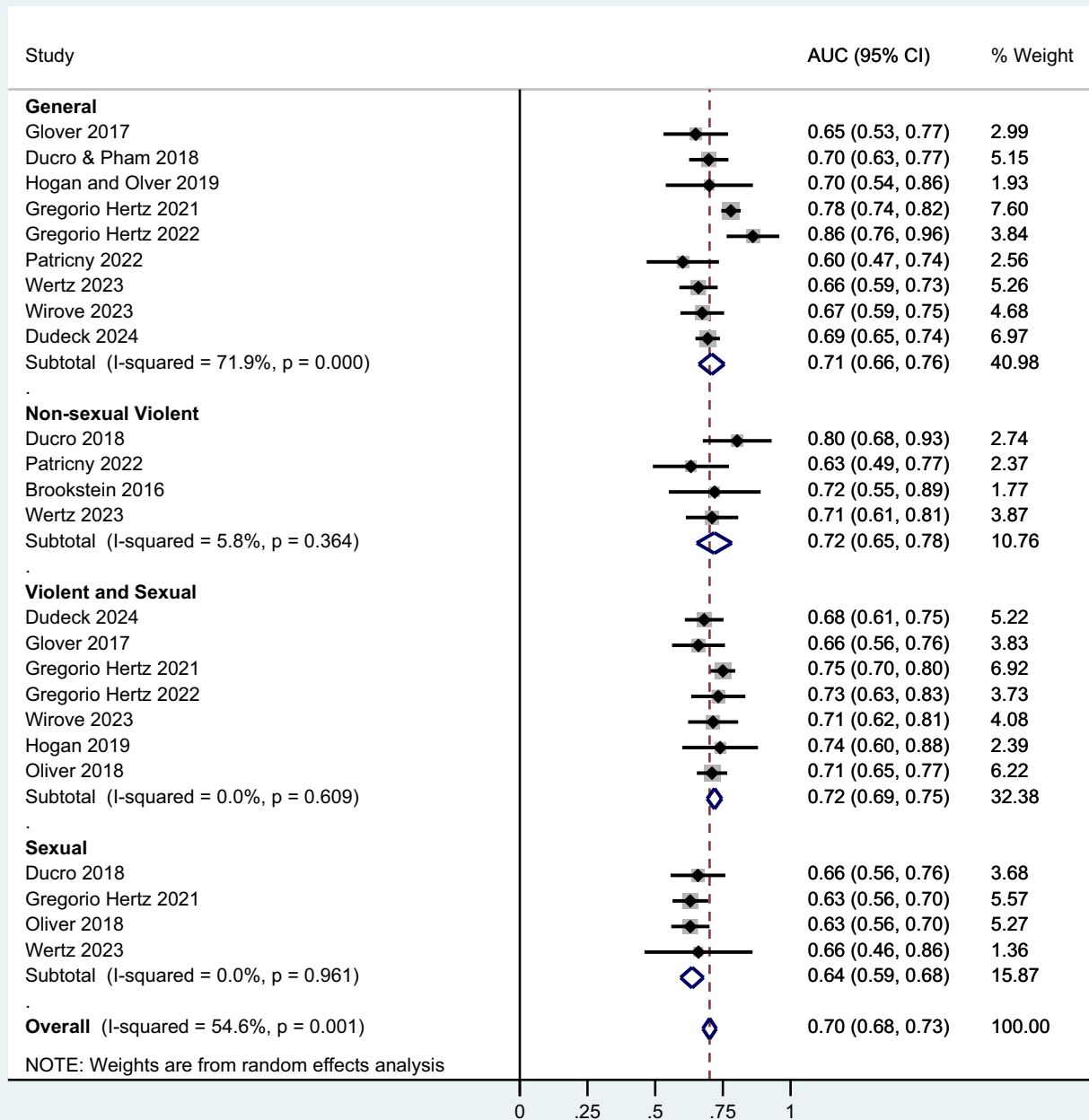


Figure 4. Meta-analysis of the predictive validity of the VRAG-R for general, nonsexual violent, violent and sexual, and sexual recidivism for more than five years of follow-up. AUC = area under the curve; CI = confidence interval.

commonly used risk assessment tools in forensic psychiatry evaluations. Despite the similar predictive validity observed, it remains pertinent to interpret these findings within the broader context of evidence that continues to advocate for the integration of actuarial risk assessment tools with clinical judgment.⁴⁴ This suggestion is based on numerous studies that, although

clinical judgment alone underperforms in predictive accuracy when directly compared with actuarial risk assessment, the combination of the two approaches yields the highest predictive accuracy in comparison to either method in isolation.⁴⁵⁻⁴⁷

Furthermore, although actuarial risk assessment typically reduces sources of bias and inconsistencies

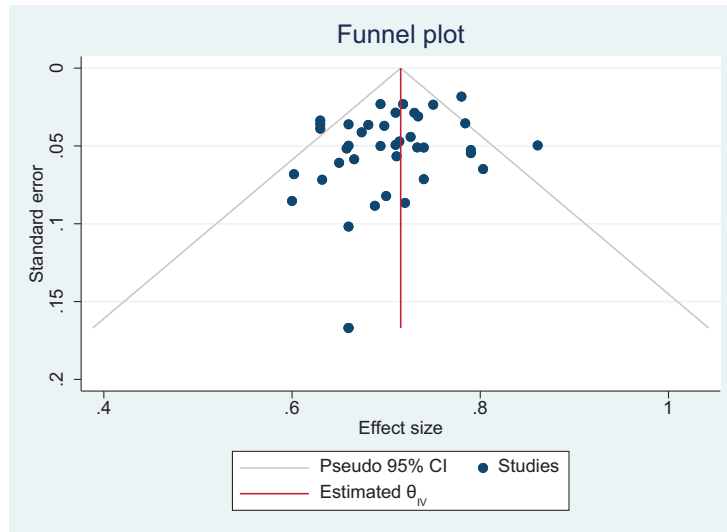


Figure 5. Funnel plot assessing publication bias. CI = confidence interval.

among and within assessors and removes variability introduced by irrelevant factors, there are many areas relevant to risk assessment that are not well captured by existing models or tools.⁴⁴ One of the earliest and most persistent critiques of actuarial risk assessment tools, however, is their limited consideration of the diversity in the populations they aim to assess. This concern was clearly illustrated in our findings, where only one out of 15 studies focused solely on females, the majority of the population was White, and only one study was completely in a country with a majority Black, Indigenous, or people of color (BIPOC) population. This risk, although contentious, has been observed in a study involving 25,980 participants, with the validity of multiple actuarial risk assessment tools found to have greater predictive validity when used in populations similar to their original validation samples, which are majority middle-aged White men.¹⁰ Finally, there were no validation studies completed in the United States of America. This may have been related to the poor generalizability of the original VRAG in American populations as well as the abundance of other risk assessment tools that are more commonly used.^{48,49}

Limitations of the Review of Literature

As previously mentioned, the population in the studies was largely homogenous, which raises concerns about the generalizability of the VRAG-R across gender identities, ages, and ethnicities. Further exacerbating this limitation was our decision to only include

English language studies, which could exclude relevant findings from non-English-speaking populations.

Examining the included studies, there were several methodological limitations that could be addressed in future research. First, there was significant variability in the definition of violent recidivism, particularly regarding whether sexual offenses were accounted for in this measure and which types of sexual offenses were included. Although we separated our pooled estimates for violent recidivism by its inclusion or exclusion of sexual offenses, this heterogeneity may still affect the validity of our meta-analytic conclusions and limit the comparability of our estimates with other risk assessment tools. Second, although all the included studies reported predictive validity, almost none reported calibration data. Although this is not a problem isolated to our meta-analysis, it continues to represent concerns with how well predicted risks align with actual outcomes. Furthermore, although most studies reported limited demographic data, more comprehensive reporting of variables such as age, race or ethnicity, gender, mental health diagnoses, and socioeconomic status would have provided more opportunity for subgroup analyses to examine the predictive validity of the VRAG-R in different populations.

Finally, there is an urgent need for more diversity and inclusion in research participation and recruitment. Given that Indigenous and other ethnocultural groups are disproportionately represented within North American criminal justice systems, it is problematic that most recidivism risk assessment tools have been developed and validated in predominantly White

populations.⁵⁰ Different ethnocultural groups may have different offense patterns and risk factors for recidivism that are unaccounted for in current risk assessment tools.⁵¹ Apart from the possibility of inaccurate risk assessment, it also reduces opportunities to identify risk factors that could be modifiable to reduce an individual's risk and allow for culturally competent interventions. More research is also needed to examine the tool's predictive validity across the lifespan and in different genders. Of the three studies focused on youth, only one examined sexual offending, whereas another reported that the VRAG-R demonstrated comparatively weaker predictive validity. Similarly, few studies had female participants, and the only study that focused solely on females demonstrated only moderate predictive validity. Thus, longstanding concerns remain regarding the generalizability of risk assessment tools that have only been validated in a homogenous sample.

Given the comparably lower predictive validity of the VRAG-R for sexual recidivism, more data on sexual offenses is needed to assess the tool's performance in this domain. This limitation also extends to the original VRAG-R validation study, where violent recidivism was inclusive of sexual assault.⁶ Moreover, as the VRAG-R was introduced as a replacement of the SORAG, more research is needed to examine the performance of the VRAG-R in predicting violent and sexual recidivism in sexual offenders, which was the original purpose behind the creation of the SORAG. Although this approach is in keeping with our increasing recognition of sexual violence as a distinct form of violence, it also addresses concerns raised by the even lower AUC values observed for sexual recidivism in sexual offenders compared with our pooled estimates for sexual recidivism in all offenders.

Finally, the mild asymmetry on visual inspection of our funnel plot suggests the possibility of publication bias, whereby smaller studies with negative findings on the VRAG-R were not included, which may inaccurately inflate the pooled estimated AUC values across all recidivism categories.

Conclusion

To our knowledge, this is the first systematic review and meta-analysis examining the predictive validity of the VRAG-R across general, violent, and sexual recidivism since its introduction in 2013. Similar to its predecessor, the VRAG-R demonstrates acceptable to good discriminative accuracy for general

and violent recidivism, although its performance in predicting sexual recidivism is more modest. These findings underscore its continued clinical relevance but also its limitations, including concerns regarding its generalizability.

As the validation samples remained demographically homogenous, their generalizability and applicability across culturally and contextually diverse populations remain ambiguous.

Future research in this area would benefit from methodological improvements, such as the reporting of both discrimination and calibration metrics. In addition, future research should prioritize examining the tools across diverse populations, including different age groups, women, racialized groups, and non-English-speaking populations.

Although the predictive validity of the VRAG-R for sexual recidivism is comparable with that of the other commonly used tools, it showed only modest ability to discriminate between those who do and do not reoffend, highlighting the need for more research in the variety of factors that contribute to sexual recidivism. Furthermore, future progress may be most impactful if focused on rigorous and inclusive validation practices instead of incremental refinement of existing tools designed for the same homogenous populations.

References

1. Wertz M, Schobel S, Schiltz K, Rettenberger M. A comparison of the predictive accuracy of structured and unstructured risk assessment methods for the prediction of recidivism in individuals convicted of sexual and violent offense. *Psychol Assess*. 2023 Feb; 35(2):152–64
2. Monahan J. The clinical prediction of violent behavior [Internet]; 1981. Available from: <https://www.ojp.gov/ncjrs/virtual-library/abstracts/clinical-prediction-violent-behavior>. Accessed November 10, 2025
3. Alba AC, Agoritsas T, Walsh M, *et al*. Discrimination and calibration of clinical prediction models: Users' guides to the medical literature. *JAMA*. 2017 Oct; 318(14):1377–84
4. Rice ME, Harris GT. Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r*. *Law & Hum Behav*. 2005; 29(5):615–20
5. Harris GT, Rice ME, Quinsey VL. Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Crim Just & Behav*. 1993; 20(4):315–35
6. Rice ME, Harris GT, Lang C. Validation of and revision to the VRAG and SORAG: The Violence Risk Appraisal Guide-Revised (VRAG-R). *Psychol Assess*. 2013; 25(3):951–65
7. Helmus LM, Quinsey VL. Predicting violent reoffending with the VRAG-R: Overview, controversies, and future directions for actuarial risk scales. In Wormith JS, Craig L, Hogue TE, editors. *The Wiley Handbook of What Works in Violence Risk Management: Theory, Research, and Practice*, First Edition. Chichester, U.K.: Wiley Blackwell; 2020. p. 119–44

8. Hanson RK, Thornton D. Static-99: Improving actuarial risk assessments for sex offenders [Internet]; 1999. Available from: <https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/sttc-mprvng-actrl/sttc-mprvng-actrl-eng.pdf>. Accessed October 28, 2025
9. Douglas KS, Hart SD, Webster CD, et al. Historical-Clinical-Risk Management-20, Version 3 (HCR-20^{V3}): Development and overview. *Int J Forensic Ment Health*. 2014; 13(2):93–108
10. Singh JP, Grann M, Fazel S. A comparative study of violence risk assessment tools: A systematic review and meta-regression analysis of 68 studies involving 25,980 participants. *Clin Psychol Rev*. 2011; 31(3):499–513
11. Douglas T, Pugh J, Singh I, et al. Risk assessment tools in criminal justice and forensic psychiatry: The need for better data. *Eur Psychiatry*. 2017; 42:134–7
12. Hart SD. Psychopathy checklists. In Jamieson A, Moenssens AA, editors. *Wiley Encyclopedia of Forensic Science*. Hoboken, NJ: John Wiley and Sons. 2009
13. Quinsey VL, Harris GT, Rice ME, et al. *Violent Offenders: Appraising and Managing Risk*. Washington, DC: American Psychological Association; 2015
14. Harris GT, Rice ME. Actuarial assessment of risk among sex offenders. *Ann N Y Acad Sci*. 2003; 989(1):198–210
15. Ramshaw L, Chatterjee S, Glancy G, Wilkie T. Canadian guidelines for forensic psychiatry assessment and report writing: General principles; Violence risk assessment [Internet]; 2021. Available from: <https://www.capl-acpd.org/wp-content/uploads/2022/06/01-Guidelines-FIN-EN-Web.pdf>. Accessed September 28, 2025
16. Ogonah MGT, Seyedsalehi A, Whiting D, Fazel S. Violence risk assessment instruments in forensic psychiatric populations: A systematic review and meta-analysis. *Lancet Psychiatry*. 2023; 10(10):780–9
17. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*. 2021; 372:n71
18. Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: The CHARMS checklist. *PLoS Med*. 2014; 11(10):e1001744
19. Fernandez-Felix BM, López-Alcalde J, Roqué M, et al. CHARMS and PROBAST at your fingertips: A template for data extraction and risk of bias assessment in systematic reviews of predictive models. *BMC Med Res Methodol*. 2023 Feb; 23(1):44
20. Veritas Health Innovation. *Covidence Systematic Review Software*. Melbourne, Victoria, Australia: Veritas Health Innovation. 2024
21. Hanson RK. Assessing the calibration of actuarial risk scales: A primer on the E/O index. *Crim Just & Behav*. 2017; 44(1):26–39
22. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986 Sep; 7(3):177–88
23. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. *BMJ*. 1997; 315(7109):629–34
24. Dudeck M, Streb J, Mayer J, et al. Evaluation of whether commonly used risk assessment tools are applicable to women in forensic psychiatric institutions. *Compr Psychiatry*. 2024; 135: 152528
25. Glover AJJ, Churcher FP, Gray AL, et al. A cross-validation of the Violence Risk Appraisal Guide-Revised (VRAG-R) within a correctional sample. *Law & Hum Behav*. 2017; 41(6):507–18
26. Gregório Hertz P, Eher R, Etzler S, Rettenberger M. Cross-validation of the revised version of the Violence Risk Appraisal Guide (VRAG-R) in a sample of individuals convicted of sexual offenses. *Sex Abuse*. 2021; 33(1):63–87
27. Gregório Hertz P, Müller M, Barra S, et al. The predictive and incremental validity of ADHD beyond the VRAG-R in a high-risk sample of young offenders. *Eur Arch Psychiatry Clin Neurosci*. 2022; 272(8):1469–79
28. Wirove RL, Olver ME, Haag A. Discrimination and calibration properties of the Violence Risk Appraisal Guide-Revised in a not criminally responsible provincial population. *Assessment*. 2023; 30(5):1672–87
29. Olver ME, Sewall LA. Cross-validation of the discrimination and calibration properties of the VRAG-R in a treated sexual offender sample. *Crim Just & Behav*. 2018; 45(6):741–61
30. Barra S, Bessler C, Landolt MA, Aebi M. Testing the validity of criminal risk assessment tools in sexually abusive youth. *Psychol Assess*. 2018; 30(11):1430–43
31. Ducro C, Pham TH. Convergent, discriminant and predictive validity of two instruments to assess recidivism risk among released individuals who have sexually offended: The SORAG and the VRAG-R. *Int J Risk Recover*. 2022; 5(1):14–28
32. Hogan NR, Olver ME. Assessing risk for aggression in forensic psychiatric inpatients: An examination of five measures. *Law & Hum Behav*. 2016; 40(3):233–43
33. Hogan NR, Olver ME. A prospective examination of the predictive validity of five structured instruments for inpatient violence in a secure forensic hospital. *Int J Forensic Ment Health*. 2018; 17(2):122–32
34. Patricny N, Haag AM, Pei JR. Resistance to antisocial peers in adolescents found not criminally responsible on account of mental disorder: Predictive and incremental validity with the VRAG-R. *Crim Just & Behav*. 2022; 49(5):681–99
35. Brookstein DM. The predictive validity of the Historical Clinical Risk Management-20 Version 3 (HCR-20V3) and the Violence Risk Appraisal Guide-Revised (VRAG-R) [unpublished doctoral dissertation]. Melbourne, VIC, Australia: Swinburne University of Technology; 2016
36. Surjanovic N, Loughin TM. Improving the Hosmer-Lemeshow goodness-of-fit test in large models with replicated Bernoulli trials. *J Appl Stat*. 2024; 51(7):1399–411
37. Nunes K, Firestone P, Bradford J, et al. A comparison of modified versions of the Static-99 and the Sex Offender Risk Appraisal Guide. *Sex Abuse*. 2002; 14(3):253–69
38. Rettenberger M, Rice ME, Harris GT, Eher R. Actuarial risk assessment of sexual offenders: The psychometric properties of the Sex Offender Risk Appraisal Guide (SORAG). *Psychol Assess*. 2017; 29(6):624–38
39. Brankley AE, Babchishin KM, Hanson RK. STABLE-2007 demonstrates predictive and incremental validity in assessing risk-relevant propensities for sexual offending: A meta-analysis. *Sex Abuse*. 2021; 33(1):34–62
40. Ducro C, Pham T. Evaluation of the SORAG and the Static-99 on Belgian sex offenders committed to a forensic facility. *Sex Abuse*. 2006; 18(1):15–26
41. de Vogel V, de Ruiter C, van Beek D, Mead G. Predictive validity of the SVR-20. *Law & Hum Behav*. 2004; 28(3):235–51
42. Tsao IT, Chu CM. An exploratory study of recidivism risk assessment instruments for individuals convicted of sexual offenses in Singapore. *Sex Abuse*. 2021; 33(2):157–75
43. Rettenberger M, Boer DP, Eher R. The predictive accuracy of risk factors in the Sexual Violence Risk-20 (SVR-20). *Crim Just & Behav*. 2011; 38(10):1009–27
44. Dawes RM, Faust D, Meehl PE. Clinical versus actuarial judgment. *Science*. 1989; 243(4899):1668–74
45. Serin RC, Mailloux DL, Branch R, Hucker S. The utility of clinical and actuarial risk assessments for offenders in pre-release psychiatric decision-making [Internet]; 2000. Available from: https://publications.gc.ca/collections/collection_2010/scc-csc/PS83-3-95-eng.pdf. Accessed October 30, 2025

A Systematic Review of the Predictive Validity of the VRAG-R

46. Doyle M, Dolan M. Violence risk assessment: Combining actuarial and clinical information to structure clinical judgements for the formulation and management of risk. *J Psychiatr Ment Health Nurs*. 2002; 9(6):649–57
47. Viljoen JL, Goossens I, Monjazebe S, *et al*. Are risk assessment tools more accurate than unstructured judgments in predicting violent, any, and sexual offending? A meta-analysis of direct comparison studies. *Behav Sci & L*. 2025; 43(1):75–113
48. Desmarais SL, Johnson KL, Singh JP. Performance of recidivism risk assessment instruments in U.S. correctional settings. *Psychol Serv*. 2016; 13(3):206–22
49. Mills JF, Jones MN, Kroner DG. Examination of the generalizability of the LSI-R and VRAG probability bins. *Crim Just & Behav*. 2005; 32(5):565–85
50. Robinson P, Small T, Chen A, Irving M. Over-representation of Indigenous persons in adult provincial custody, 2019/2020 and 2020/2021 [Internet]; 2023. Available from: <https://www150.statcan.gc.ca/n1/pub/85-002-x/2023001/article/00004-eng.htm>. Accessed October 29, 2025
51. van der Put C, Stams GJ, Dekovi M, *et al*. Ethnic differences in offense patterns and the prevalence and impact of risk factors for recidivism. *Int'l Crim Just Rev*. 2013; 23(2):113–31