Applying a Forensic Actuarial Assessment (the Violence Risk Appraisal Guide) to Nonforensic Patients

GRANT T. HARRIS  
MARNIE E. RICE  
Mental Health Centre Penetanguishene  
JOSEPH A. CAMILLERI  
University of Saskatchewan

The actuarial Violence Risk Appraisal Guide (VRAG) was developed for male offenders where it has shown excellent replicability in many new forensic samples using officially recorded outcomes. Clinicians also make decisions, however, about the risk of interpersonal violence posed by nonforensic psychiatric patients of both sexes. Could an actuarial risk assessment developed for male forensic populations be used for a broader clientele? We modified the VRAG to permit evaluation using data from the MacArthur Violence Risk Assessment Study that included nonforensic male and female patients and primarily self-reported violence. The modified VRAG yielded a large effect size in the prediction of dichotomous postdischarge severe violence over 20 and 50 weeks. Accuracy of VRAG predictions was unrelated to sex. The results provide evidence about the robustness of comprehensive actuarial risk assessments and the generality of the personal factors that underlie violent behavior.

Keywords: violence risk assessment; actuarial; prediction; civil psychiatric patients; MacArthur Violence Risk Assessment Study

The past decade has seen a proliferation of instruments designed to predict violence. The majority of these instruments have been designed to predict criminal violence among offenders with mental disorders or criminal offenders, especially sex offenders (e.g., Barbaree, Seto, Langton, & Peacock,

Authors’ Note: We wish to acknowledge the MacArthur Foundation and the MacArthur Risk Network members for generously allowing researchers to use data from the MacArthur Violence Risk Assessment Study. We also acknowledge support of the Ontario Ministry of Health and Long-Term Care, as well as Zoe Hilton, John Monahan, and Vern Quinsey for helpful comments on earlier versions of this article. Address correspondence to: gharris@mhcp.on.ca.
One of the earliest, most accurate, and most replicated of these instruments is the Violence Risk Appraisal Guide (VRAG; Harris, Rice, & Quinsey, 1993). The VRAG is a purely actuarial instrument developed on a sample of 618 male offenders with mental disorders and subsequently convicted offenders assessed in a maximum-security psychiatric hospital in Ontario, Canada. The instrument was shown to yield a high degree of accuracy (relative operating characteristic [ROC] area of 0.76) for the development sample in the prediction of a subsequent criminal act of violence over an average time at risk of 7 years. The ability of the VRAG to predict subsequent criminal violence among criminal offenders with mental disorders has been replicated in more than 25 studies in at least five different countries (see www.mhcp-research.com/ragreps for a complete list of replications).

The VRAG has been shown to predict future criminal violence over mean follow-up periods ranging from 15 months (Quinsey, Book & Skilling, in press) to 10 years (Rice & Harris, 1995) and in samples with base rates of violent recidivism ranging from 22% (Rice & Harris, 2002) to 57% (Rice & Harris, 1995). It has also been shown to predict time until the first violent reoffense and severity of violent offense (e.g., Harris, Rice, & Cormier, 2002; Harris, Rice, Quinsey, Lalumière, Boer, & Lang, 2003). In addition to violent recidivism, the VRAG has been shown to exhibit predictive validity for the outcomes of general criminal recidivism (Loza, Villeneuve, & Loza-Fanous, 2002; Nugent, 2001), institutional misconduct (Kroner & Mills, 2001; McBride, 1999), institutional violence (Nadeau, Nadeau, Smiley, & McHattie, 1999; Nichols, Vincent, Whittemore, & Ogloff, 1999), and sexual recidivism (Barbaree et al., 2001; Harris et al., 2003). The predictive accuracy of the VRAG has been shown to be greater in studies where there is little or no variance in follow-up time, scoring reliability is high, and no VRAG items are omitted or approximated (Harris & Rice, 2003). Under such optimal conditions, the predictive accuracy of the VRAG in predicting violent recidivism has been shown to exceed an ROC area of 0.85 (see also Dempster, Hart, & Boer, 2002; Pham, 2002).

The VRAG was developed exclusively on a sample of male offenders. To date, there has been only one study that has reported on its application to women. In that study (Harris et al., 2002), it was found that the instrument failed to predict criminal violence among women. However, the number of women (all of whom were forensic psychiatric patients) was small (N = 59), their risk scores were low, and the range of scores among the women was restricted compared to those of the men.
The MacArthur Violence Risk Assessment Study (Monahan et al., 2001) included more than 1,000 nonforensic inpatients (58% of whom were admitted voluntarily) from three public hospitals in the United States. Approximately 40% were women. The measurement of study variables entailed remarkable efforts to ensure reliability and consistency. The MacArthur Violence Risk Assessment Study investigators generously made their data available to other researchers wishing to use them for other studies. The availability of these data suggested an ideal opportunity to examine the generalizability of the VRAG by testing it on a large nonforensic sample of male and female psychiatric patients, using a very different outcome measure (i.e., almost entirely, those violent acts reported by the patient or a collateral), and using a shorter follow-up period (20 weeks) than previously reported.

METHOD

Archival Data

The data for the current study were downloaded from the MacArthur Violence Risk Assessment Study Web site (www.macarthur.virginia.edu). The posted data came from all patients who consented (N = 1136) to participate in the MacArthur project. The data were downloaded as SPSS files, as were the coding manuals and scoring criteria. The details of the MacArthur Violence Risk Assessment Study methods have been extensively described in book form (Monahan et al., 2001). In summary, the participants were voluntary and civilly committed inpatients from three public hospitals in Pittsburgh, Pennsylvania; Kansas City, Missouri; and Worcester, Massachusetts. All were between ages 18 and 40 years and spoke English. Selection was limited to White, African American, and Hispanic racial groups, and some diagnoses were excluded (primarily anxiety, dementia, sexual, impulse-control, eating and sleeping disorders, and disorders usually first diagnosed in juveniles).

The measurement of the study variables entailed strenuous efforts to ensure reliability and consistency. Trained interviewers conducted lengthy predischarge interview sessions (supplemented by record reviews) to gather the personal, historical, social, and clinical characteristics, of which there were 134. Postdischarge outcome was assessed primarily by interviewing participants and collaterals every 10 weeks for 50 weeks using a modification of the Conflict Tactics Scale (CTS; Straus, 1979) that yielded estimates for nine different aggressive behaviors ranging from pushing to using a weapon on someone, and scores for minor and severe violence. Arrests and rehos-
pitalizations for acts of violence were also recorded; however, the MacArthur researchers found that violence detected by these official sources was very rare (e.g., less than 2% for arrests in the first 20 weeks). Published research on the MacArthur Violence Risk Assessment Study has primarily concerned outcomes that could be assessed only in the first 20 weeks (Monahan et al., 2001). Despite ambitious efforts to contact them through past addresses and contacts, not every participant could be located for each interview; however, the MacArthur investigators have concluded that patients lost to follow-up via this form of attrition were not different in the prevalence of subsequent violence than those not lost (Steadman et al., 1998). This suggests that the current approach yielded fair estimates of the base rate of serious violence and predictive accuracy.

Coding of VRAG Items and Outcome

In most cases, some minor modifications of the VRAG items were necessary to use the MacArthur archive. All decisions about the best approximations to the VRAG items from variables in the MacArthur data set were made prior to any analyses using outcome data. Table 1 shows the operational definition of the original VRAG item on one hand and the MacArthur analog on the other. One VRAG item, failure on a prior conditional release, could not be approximated from the MacArthur data, and another, worst injury to a victim in the index offense, was almost always (92%) missing in the MacArthur archive; these two items were dropped. Clearly, however, the MacArthur analogs capture the manifest content of the remaining 10 VRAG items.

The VRAG was designed to predict whether an offender would be charged or convicted for at least one subsequent act of criminal violence. Although charges for criminally violent acts was one of the outcome measures used in the MacArthur study, the base rate for such acts was very low. Thus, we decided that the most appropriate primary outcome variable in the current study was the presence of at least one act of severe violence on the CTS, which was the next closest analog to the outcome the VRAG was designed to predict. As well, we confined our analyses to those participants who were available for both of the first two follow-up interviews (n = 741), 26% of whom were recorded with one or more instances of serious violence. As in all the reports of the MacArthur study (Monahan et al., 2000; Monahan et al., 2001; Steadman et al., 2000), we used the 20-week follow-up as our primary follow-up period. For subsidiary analyses, we also computed the total number of seriously violent incidents and worst and total amount of injury caused, and we created our own scale for the total severity of all the reported posthospitalization violence by adding up the CTS item scores, (i.e.,
<table>
<thead>
<tr>
<th>Item Name</th>
<th>Standard (Quinsey, Harris, Rice, &amp; Cormier, 1998)</th>
<th>Adaptation</th>
<th>$rv$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Lived with parents to age 16</td>
<td>Lived with both biological parents to age 16 (except for death of a parent)</td>
<td>Lived with both biological parents until at least 15 years old</td>
<td>$-0.13$</td>
</tr>
<tr>
<td>2. Elementary school maladjustment</td>
<td>Rated from 0 = no problems to 4 = severe behavior or attendance problems</td>
<td>One point for each of hyperactivity/school difficulties, starting fights, use weapons in a fight, hurt others not in a fight; all before age 15</td>
<td>$0.21$</td>
</tr>
<tr>
<td>3. Alcohol problems</td>
<td>One point for each of parental alcoholism, teenage alcohol problem, adult alcohol problem, alcohol involved in prior offense, alcohol involved in index offense</td>
<td>One point for each of alcohol abuse diagnosis, excessive maternal alcohol, excessive paternal alcohol, drinking just prior to index admission, current alcohol abuse, lifetime alcohol abuse</td>
<td>$0.09$</td>
</tr>
<tr>
<td>4. Marital status</td>
<td>Ever married or lived common law (i.e., lived together at least 6 months)</td>
<td>Ever married or living with someone as though you were married</td>
<td>ns</td>
</tr>
<tr>
<td>5. Nonviolent offense history</td>
<td>Cormier Lang score for frequency and severity of nonviolent offenses</td>
<td>Prior arrests for property offenses</td>
<td>$0.11$</td>
</tr>
<tr>
<td>6. Age</td>
<td>Age at index offense</td>
<td>Age at index admission</td>
<td>$-0.06$</td>
</tr>
<tr>
<td>7. Victim sex</td>
<td>Any female index offense victim</td>
<td>Any female victim recorded for admitting incident</td>
<td>ns</td>
</tr>
<tr>
<td>8. Personality disorder</td>
<td>Meet DSM-II criteria for any personality disorder</td>
<td>Received baseline discharge diagnosis for personality disorder</td>
<td>$0.13$</td>
</tr>
<tr>
<td>9. Schizophrenia</td>
<td>Meet DSM-III criteria for schizophrenia</td>
<td>Received baseline discharge diagnosis of schizophrenia</td>
<td>$-0.10$</td>
</tr>
<tr>
<td>10. Psychopathy</td>
<td>20-item Revised Hare Psychopathy Checklist</td>
<td>12-item Screening Version of the Hare Psychopathy Checklist</td>
<td>$0.35$</td>
</tr>
</tbody>
</table>
1 for pushing up to 9 for using a weapon on someone) for each instance of interpersonal aggression reported in the first 20 weeks and for the full 50-week period.

RESULTS

The distribution of the modified VRAG scores was similar to that of the standard VRAG score using official records in offender populations, mean = −.92 (SD = 7.43) range −23 to 24, skewness = .16 compared to a mean of .91 (SD = 12.9), range −27 to +35, and skewness = .19 for the unmodified 12-item VRAG in our original development sample (Harris, Rice, & Quinsey, 1993). Most important, the modified and shortened VRAG yielded a large effect (though significantly smaller than achieved in its original construction, Harris, Rice, & Quinsey, 1993) in predicting subsequent seriously violent behavior, ROC area = .72, 95% CI (confidence interval) ± .04. The 552 patients without an act of serious violence had a mean modified VRAG score = −2.41 (SD = 7.1) 95% CI ± .59 compared to 3.41 (SD = 6.8) 95% CI ± .97 for the 189 patients who were subsequently seriously violent. The accuracy of the modified VRAG was unrelated to sex: ROC area = .71, 95% CI ± .06 for the 423 men and .73, 95% CI ± .06 for the 318 women, and the two sexes exhibited similar rates of violence, 27% versus 23%, respectively (Monahan et al., 2001, reported that the overall rate of violence of any kind was significantly higher among men).

In subsidiary analyses, we evaluated the association between modified VRAG score and continuous measures of the severity of violent outcome. Thus, over the 20-week period, the modified VRAG significantly predicted the number of seriously violent incidents, \( r = .33, p < .0001 \); the total number of any violent incidents, \( r = .34, p < .0001 \); the most serious injury caused in a violent incident, \( r = .25, p < .0001 \); the total amount of injury caused in all violent incidents, \( r = .24, p < .0001 \); and the total CTS scores for all violent incidents, \( r = .21, p < .0001 \). As well, the modified VRAG scores predicted violence throughout the entire 50-week follow-up. Thus, the 267 patients recorded with at least one act of serious violence had larger modified VRAG scores than the 474 without serious violence in the entire 50 weeks, −2.74 (SD = 7.2) versus 2.30 (SD = 6.9), ROC area = .70, 95% CI ± .04. Modified VRAG scores were also significantly related to number of violent incidents, \( r = .33, p < .0001 \); number of serious violent incidents, \( r = .35, p < .0001 \); arrests for violence, \( r = .13, p < .0001 \); total seriousness score, \( r = .20, p < .0001 \); and most serious injury in a violent incident, \( r = .23, p < .0001 \) over the 50-week period. In fact, no test of the ability of the modified VRAG to predict...
any index of subsequent violence failed to yield a statistically significant result. The only one that was close was arrests for violence in the first 20 weeks (base rate < 2%) where the correlation was significant, but low, $r = .06$, $p < .05$, one-tailed.

**DISCUSSION**

By making minor adjustments, we were able to score the manifest content of 10 of the 12 items constituting the VRAG. The resulting modified VRAG scores yielded a large effect size in predicting the occurrence of subsequent serious violence within a 20-week follow-up. In addition, these modified VRAG scores predicted the number of subsequent violent incidents of all types and also predicted the overall severity of all subsequent violent behavior, in the 20-week follow-up, and in the full 50-week period. The effect obtained in the current study should be considered conservative because several items had to be approximated, and two were missing. For comparison purposes, the ROC area using only the 10 items (or their approximations) used in the current study calculated for a recently completed prospective replication of the VRAG (Harris et al., 2002) was .77 (in a constant 5-year follow-up), whereas the ROC area in that study using all 12 variables was .80. This suggests that if all original 12 items of the VRAG could have been scored without modification in the current study, the ROC area would likely have been at least 0.75.

The VRAG in the current study worked at least as well for women as it did for men. The sample of women in the MacArthur data was considerably larger than the sample of women in the only previous study, to date, that has reported an ROC area for women (Harris et al., 2002). In addition, the mean risk level of the women was not significantly lower for women than men, and the range and standard deviation of the women’s scores were not different from those of the men. Thus, the data from the current study provided a stronger test of the accuracy of the VRAG than the previous one and suggest it warrants further research among female patients and offenders.

The current study entailed some significant limitations. First, most VRAG items had to be approximated by using variables in the MacArthur database that differed (usually in small ways) from those used in the VRAG, and two were unavailable. These compromises mean that the current results probably underestimate the true performance of the VRAG, however it is unclear by how much. Subsidiary analyses suggested that had it been possible to score the entire VRAG as designed, its performance would have been indistinguishable from its many other true replications (e.g., Harris et al., 2002;
Harris et al., 2003; Rice & Harris, 1997). A second limitation of the current study was inherent in the original MacArthur methodology—one of the predictor variables, the Psychopathy Checklist, was scored during the first 20 weeks of the follow-up period (Skeem, Monahan, & Mulvey, 2002). This confound raises the possibility that interviewers’ knowledge of the outcome (self-reported aggression) might have inadvertently biased psychopathy ratings or vice versa. Nevertheless, the robust ability of the Psychopathy Checklist to predict violence in a wide variety of offender populations (e.g., Salekin, Rogers, & Sewell, 1996) gives us confidence that the MacArthur results pertaining to the Psychopathy Checklist will replicate when tested without this methodological confound.

The performance of the VRAG in the current study supports the findings of previous meta-analyses of the predictors of violence (Bonta, Law, & Hanson, 1998; Gendreau, Little, & Groggin, 1996; Hanson & Bussière, 1998) that have suggested that the predictors of violence are similar across a wide variety of populations. Furthermore, these results are consistent with those of other studies that have shown that the clinical problems and clinical needs of forensic and nonforensic patient are similar (Harris, Hilton, & Rice, 1993; Seto, Harris, & Rice, in press). The results of the current study also suggest generality across different measures of violent outcome, and across a wide range of follow-up times. The current results also have implications for the understanding of the causes of violence. Even though few of the present participants would have qualified for a diagnosis of psychopathy, the personality traits associated with psychopathy are among the most important causes of human aggression. Conversely, although most of the present participants would qualify for a serious Axis I psychiatric disorder, only a few of the many tested positive psychotic symptoms were positively associated with subsequent violence (Monahan et al., 2001). The MacArthur Violence Risk Assessment data (Monahan et al., 2001) provide no empirical support for the idea that psychotic symptoms are important causes of aggression even among nonforensic psychiatric patients being considered for discharge.

Our results, together with those of other analyses of the MacArthur data (Monahan et al., 2000; Monahan et al., 2001; Steadman et al., 2000) seriously challenge the notion that positive, active symptoms of psychosis (i.e., hallucinations, delusions, formal thought disorder) are useful risk factors (for postdischarge violence) among persons who are institutionalized and suffering from serious mental illness. The current pattern of results suggests only three possibilities: active psychotic symptoms are trivial causes of aggression; pharmacological treatment reduces symptoms and standard risk assessment factors (such as the VRAG items) predict medication response or noncompliance; or standard risk assessment factors (such as the VRAG
items) predict aggression but only among those who are unresponsive to or noncompliant with pharmacological treatment; or some combination of these. The current results, and the reports from the original MacArthur investigators provide support for the use of actuarial risk assessment in psychiatric populations and yield no empirical support for adjusting actuarial scores because of psychotic symptoms in patients’ prerelease clinical presentations.

The evidence favors actuarial assessments to assess the risk of violence among persons who are institutionalized and being considered for release. The iterative classification tree yielded an ROC area of .82, higher than that of the modified VRAG. Nevertheless, the predictive accuracy of both instruments was high according to conventional standards (Cohen, 1992). Assuming that cross-validation studies continue to show superiority of the iterative classification tree (Monahan et al., 2001) it would be the instrument of choice for nonforensic psychiatric patients. However, the accuracy of the VRAG gives users a choice between two instruments with good predictive validity. Decision makers can be expected to select one based on relative accuracy, costs, convenience, and the availability of individual items.

Because ROC statistics are insensitive to variation in the base rate of the outcome variable, their use to evaluate accuracy leaves one question unanswered. Do forensic psychiatric patients represent different overall levels of risk to public safety than do nonforensic patients? Because outcome variables have been operationalized differently, the current data do not permit an easy comparison with studies of the VRAG in forensic samples. The mean VRAG score for the current data was similar to that observed in a forensic sample (Harris et al., 2002) suggesting that overall risk is similar when outcomes are defined in the same way. However, in forensic patients (Harris et al., 2002) the base rate of arrest or rehospitalization for violence was approximately 10% in the 1st year of opportunity, whereas the rate of arrest for violent offenses in the current MacArthur data was less than 4% in 50 weeks. It is unclear, however, how much sex differences in risk and variations in police arrest and charging practices might affect these rates. Further research is required to evaluate whether the forensic-nonforensic distinction by itself has any impact on the overall risk to public safety represented by psychiatric patients.

CONCLUSIONS

That the VRAG worked nearly as well, despite modifications, in a sample of patients quite different from the development sample supports the conclusion that the predictors and causes of violence are quite general rather than
specific to particular populations. The current results support the conclusion that the operational definition of violence, duration of follow-up, sex, and the forensic-nonforensic distinction are all irrelevant to the assessment as to which patients in any sample are at risk for subsequent violence. All these findings imply that actuarial risk assessments generalize to new populations, follow-up durations, and measures of violence better than had been first thought. The MacArthur Violence Risk Assessment Study (and our own research, Harris et al., 2002) indicates that modifying or setting aside actuarial scores based on the severity of active psychotic symptoms reduces validity. Moreover, the fact that psychopathy is such a robust predictor of violence across populations suggests that personality traits associated with psychopathy must be among its most important causes.

REFERENCES


Grant T. Harris is director of research at the Mental Health Centre, Penetanguishene, Ontario, Canada. He is also an adjunct associate professor of psychology at Queen’s University at Kingston, and adjunct associate professor of psychiatry at the University of Toronto. He obtained a Ph.D. in experimental psychology from McMaster University, Hamilton, Ontario, Canada. He, together with Dr. Rice, has been awarded several research grants and has conducted extensive scientific research on violent and criminal behavior, psychopathy, and sexual aggression. His recent research has included the development of the actuarial Ontario Domestic Assault Risk Assessment.

Marnie E. Rice is the scientific director of the Centre for the Study of Aggression and Mental Disorder associated with McMaster University and the Mental Health Centre Penetanguishene. She is professor of psychiatry and behavioural neurosciences at McMaster University, professor of psychiatry at University of Toronto, and associate professor of psychology at Queen’s University. She has won the American Psychological Association’s Distinguished Contributions to Research in Public Policy Award and the Saleem Shah Memorial Lecture Award from the U.S. National Association of State Mental Health Directors’ Forensic Division. She was recently inducted as a Fellow of the Royal Society of Canada.

Joseph A. Camilleri received an honours bachelor’s degree in psychology from McMaster University and is a graduate student at the University of Saskatchewan. His research focuses on the integration of forensic, evolutionary, and applied social psychology. Specifically, he is studying spousal rape, the treatment of paraphilic sexual arousal, and the statistical evaluation of clinical treatment programs for rapists. He is also the chair-elect for the Student’s Section of the Canadian Psychological Association.