Psychological Assessment
2011, Vol. 23, No. 1, 174–183

1040-3590/11/$12.00 DOI: 10.1037/a0021290

Predictive and Incremental Validity of the Violence Risk Appraisal Guide Scores With Male and Female Jail Inmates

Mark E. Hastings
George Mason University and
Loudoun County Mental Health Center

Shilpa Krishnan, June P. Tangney, and Jeffrey Stuewig
George Mason University

The present study examines the predictive and incremental validity of Violence Risk Appraisal Guide scores in a sample of 328 male and 145 female jail inmates held on felony charges. Significant gender differences were observed in VRAG item and total score means, as well as in correlations between the VRAG and concurrent measures of aggression. VRAG scores significantly predicted institutional misconduct during incarceration and recidivism in the first year postrelease for male inmates but not for female inmates. In terms of incremental validity, VRAG scores predicted institutional misconduct and recidivism beyond that accounted for by psychopathy for male inmates but not for female inmates. Implications for clinical practice and future research are discussed.

Keywords: Violence Risk Appraisal Guide, validity, inmates, gender, recidivism

Supplemental materials: http://dx.doi.org/10.1037/a0021290.supp

The assessment of risk for future violence has become a staple of forensic psychological practice (Monahan, 1996). The advent of actuarial risk instruments such as the Violence Risk Appraisal Guide (VRAG; Quinsey, Harris, Rice, & Cormier, 1998) have contributed mightily to this endeavor. The VRAG is a 12-item actuarial risk assessment tool that has been validated for use in a wide variety of populations such as sex offenders (Harris et al., 2003; Langton et al., 2007), civil psychiatric patients (Harris, Rice, & Camilleri, 2004), mentally disordered (Gray, Fitzgerald, & Taylor, 2007) and non–North American offender samples (Doyle, Dolan, & McGovern, 2002; Kröner, Stadtland, & Eidt, 2007; Urbaniok, Noll, Grunewald, Steinbach, & Endrass, 2006). The ability of the VRAG to predict violent behavior among criminal and mentally disordered male inmates has been well established (Campbell, French, & Gendreau, 2009; Glover, Nicholson, Bernfeld, & Quinsey, 2002; Kroner & Mills, 2001).

Women represent one of the fasting growing population of inmates (Bureau of Justice Statistics, 2009), yet the field of violence risk assessment has not kept pace with this particular subgroup of inmates. There are no actuarial risk assessment instruments designed specifically for use with female inmates, and the validity of interpretation of VRAG scores with female inmates remains largely unknown. Female inmates share some risk factors for violence and recidivism in common with male inmates, specifically psychopathy (Loucks & Zamble, 1999; Salekin, Rogers, Ustad, & Sewell, 1998) and prior criminal history (Bonta, Pang, & Wallace-Capretta, 1995). However, research also indicates that female inmates are not simply "male inmates in drag" but rather possess unique treatment needs and pathways to violence and offending (Bonta et al., 1995). For example, Van Voorhis, Wright, Salisbury, and Bauman (2010) found that gender-responsive variables make for stronger predictive schemes among female inmates compared with gender-neutral variables alone. These researchers also found that long-established predictors of recidivism among male inmates, such as criminal thinking and antisocial associates, were much less relevant to recidivism among female inmates. Yet instruments, such as the VRAG, are often used in correctional settings despite the fact that they were not originally created, or extensively validated, for women.

Walters (2006) conducted a meta-analysis of the predictive utility of actuarial risk assessment instruments compared with self-report measures. Six samples included in the analysis were female or male–female mixed samples. Overall, the mean weighted effect size for male compared with female studies was comparable. However, there was greater variability in the range of effect sizes in the female sample studies, and none of the female samples utilized the VRAG. Campbell, French, and Gendreau's (2009) more recent meta-analysis examined the predictive utility of a broader range of risk instruments and other psychological measures, including the VRAG. The VRAG performed well in predicting institutional violence and violent recidivism. However, of the effect sizes where gender composition of the samples was

Mark E. Hastings, Department of Psychology, George Mason University, and Loudoun County Mental Health Center, Leesburg, Virginia; Shilpa Krishnan, June P. Tangney, and Jeffrey Stuewig, Department of Psychology, George Mason University.

We wish to thank all of the lab members at the Human Emotions Research Laboratory, George Mason University, and the inmates who participated in this project. We also thank James E. Hastings for his valuable comments on drafts of this article. We acknowledge the National Institute on Drug Abuse for funding the research described in this article. All opinions expressed in this article are the sole opinions of the authors and do not reflect opinions or policy of George Mason University, the Loudoun County Mental Health Center, or any other agency.

Correspondence concerning this article should be addressed to Mark E. Hastings, Department of Psychology, George Mason University, 4400 University Drive, Fairfax, VA 22030. E-mail: Mark.Hastings@loudoun.gov

reported, the vast majority were generated from male samples. It was not possible to examine gender as a moderator of predictive utility for any of the risk assessments, including the VRAG.

Only three studies to date have specifically assessed the predictive validity of VRAG scores in female samples. Harris, Rice, and Cormier (2002) reported on the predictive accuracy of the VRAG in a small sample of female forensic patients embedded within a larger mixed gender sample. They found that the VRAG was unrelated to subsequent violent behavior in the 59 female patients in the sample. However, they also noted that the base rate of violence in the female sample was low. Examination of the individual VRAG items for women revealed that the victim injury (Item 8) was inversely related to subsequent violence and that the Hare Psychopathy Checklist—Revised (PCL–R [Hare, 1991]; Item 12) was unrelated. Information on the other VRAG items was not reported and, as such, these items probably were not significantly related to subsequent violent behavior for women.

Harris et al. (2004) examined VRAG data on 423 male and 318 female civil psychiatric patients who participated in the MacArthur Violence Risk Assessment Project. A modified VRAG was used in this study owing to a lack of information needed to score several VRAG items. The abbreviated VRAG was found to predict violence as well for women (area under the curve [AUC] = .73) as for men (AUC = .71), with similar rates of violence evidenced by men and women (27% and 23%, respectively). However, when Edens, Skeem, and Douglas (2006) conducted incremental validity analyses on this same data set, they found that for the sample as a whole, modified VRAG scores contributed minimally to the prediction of violence once the Hare Psychopathy Checklist: Screening Version (PCL:SV) was controlled. No gender-specific incremental validity analyses were conducted.

Coid et al. (2009) examined the predictive validity of scores from five risk assessment instruments with both male and female offenders, including the VRAG. The VRAG was predictive of violent (AUC = .65), general (AUC = .66), and any (AUC = .66) recidivism in female offenders; however, the level of predictive accuracy was small to moderate, and incremental validity analyses were not conducted. Last, the VRAG was significantly less predictive of violent recidivism for female compared with male offenders.

Only one study to date has examined the utility of the VRAG with a female correctional sample, and no study has examined how the individual VRAG items may differ in their relationship, both concurrently and prospectively, with relevant constructs depending on gender. Also understudied is the incremental validity of VRAG scores above and beyond Hare psychopathy scores separately by gender.

The present study attempts to fill these gaps in the literature by examining the concurrent, predictive, and incremental validity of the VRAG total score as well as the individual items. We directly compare the performance of these items for men versus women. Understanding the differential relationships, if any, may help improve our use of risk appraisal procedures for women.

Method

Participants

Participants were 473 adults (328 men and 145 women) recently incarcerated at a county jail just outside Washington, DC, who were held on at least one felony charge or serving a sentence of at least 4 months. All participants were housed in medium or maximum security general population. Participants completed the Personality Assessment Inventory (PAI; Morey, 1991, 2007) using touch-screen computers that require minimal familiarity with computers (e.g., no keyboard, no mouse). In addition to presenting questionnaire items visually, the computer read each item aloud to all participants via headphones, thus accommodating participants with limited reading proficiency.

Table 1 shows the demographics of the sample. Participants ranged in age from 18.2 to 69.7 years, and there was a significant gender difference, with female inmates being on average 3.5 years older than male inmates. Education ranged from 5 to 19 years, and there was a modest but statistically significant gender difference with female inmates reporting on average nearly one more year of education than male inmates. Participant IQ (Wonderlic, 1999) ranged from 64 to 138, with no difference between men and women. PCL:SV Psychopathy scores ranged from 1 to 24. Consistent with much previous research, men scored higher than women. With regard to race, both male and female samples were diverse, but there was no gender difference.

Table 1
Demographic Characteristics and Criminal History of Jail Inmates

Demographic	Total	Male	Female	Inferential statistic
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*t* (*df*)
Age	32.8 (10.0)	31.7 (9.8)	35.2 (10.2)	3.5** (471)
Years of education	11.9 (2.1)	11.7 (2.1)	12.4 (2.1)	3.7** (468)
IQ	94.3 (13.7)	94.0 (13.5)	94.8 (14.3)	0.59 (442)
Hare Psychopathy Checklist	12.3 (4.8)	13.2 (4.8)	10.4 (4.3)	−6.2** (471)
	N (%)	*N* (%)	*N* (%)	
Demographic race				$\chi^2(3, N = 473) = 5.7$
African American	217 (46)	155 (47)	62 (43)	
Caucasian	178 (38)	113 (35)	65 (45)	
Hispanic	26 (5)	21 (6)	5 (3)	
Other/mixed	52 (11)	39 (12)	13 (9)	

* $p < .05$. ** $p < .01$.

Measures and Procedures

Shortly after assignment to the general population, eligible inmates were presented with a description of the study and asked to participate. Inmates were assured of the voluntary and confidential nature of the project. In particular, it was emphasized that the decision to participate or not would have no bearing on their status at the jail nor their release date. Of those approached, 74% agreed to participate and provided written consent. The main reason cited for nonparticipation was anticipated imminent release. There was no gender difference in rate of consent. Of those who consented, 82% ($n = 518$) remained in the jail long enough to complete portions of the assessment relevant to this report. The main reason for noncompletion was being bonded (bailed) out of jail. Of the 518, 5% ($n = 29$) were excluded because of validity concerns based primarily on two validity scales from the PAI—Inconsistency and Infrequency. Specifically, participants were dropped if either Inconsistency or Infrequency was above the recommended cutoffs (T scores of 72 and 74, respectively; Morey, 1991), and the other was considered elevated (T score above 69). Second, interviewers routinely reported any validity concerns during data collection. In cases where interviewer concerns were raised, data were further examined for response bias, response sets, elevation of other validity scales, and documentation of problems from other sessions. Finally, 16 participants (all men) were Spanish-only speakers who completed the PAI in face-to-face interviews instead of on the touch-screen computer. Because of the different mode of administration, plus the gender difference in the group, these individuals were subsequently dropped from these analyses, leaving a final sample for analysis of 473. With regard to confidentiality, interviews were conducted in the privacy of professional visiting rooms used by attorneys or in secure classrooms. In addition, data were protected by a Certificate of Confidentiality from the U.S. Department of Health and Human Services. Inmates who agreed to participate, and who completed the 4- to 6-session intake assessment, received a $15–$18 honorarium.

Hare Psychopathy Checklist: Screening Version. The PCL:SV (Hart, Cox, & Hare, 1995) was used to assess psychopathy. Completion of this 12-item checklist requires an in-depth psychosocial history interview, videotaped for coding with the participant's permission. In addition, as part of participation, inmates granted access to criminal and jail records. This collateral information was used by trained clinicians in conjunction with the interview to code the PCL:SV.

Prior to coding the PCL:SV, interviewers completed an advanced graduate course on theory, research, and assessment of psychopathy, including intensive supervised training in the administration and scoring of the PCL–R and the PCL:SV. Only those who successfully met interrater reliability criteria for both forms were cleared for coding study protocols. A randomly selected set of 54 cases were double-coded by a referent clinician (the first author) who brought to the project 15 years of professional experience in conducting forensic psychological evaluations, as well as advanced training and experience in the administration and scoring of the PCL–R and the PCL:SV. Single-measure intraclass correlations, determined with a one-way random-effects model, were .85, .79, and .85 for Part 1, Part 2, and total PCL:SV scores, respectively, showing a high degree of interrater reliability.

Violence Risk Appraisal Guide (The VRAG; Harris, Rice, & Quinsey, 1993: Quinsey et al., 1998). The VRAG is a 12-item actuarial risk assessment instrument developed with a sample of mentally disordered offenders. Similar to the PCL:SV, we assessed interrater reliability with a single-measure intraclass correlation using a one-way random-effects model. The randomly selected 52 cases showed high reliability (intraclass correlation coefficient [*ICC*] = .89).

Personality Assessment Inventory. The following scales from the PAI (Morey, 1991, 2007) were utilized in concurrent validity analyses: the Antisocial and Aggression scales, the Aggression–Physical subscale, and the Violence Potential Index. The Antisocial and Aggression subscales have been repeatedly found to be associated with institutional misconduct among male and female criminal offenders (Edens & Ruiz, 2006; Skopp, Edens, & Ruiz, 2007) and with recidivism (Boccaccini, Murrie, Hawes, Simpler, & Johnson, 2010; Walters, 2006). Although less thoroughly researched, the Violence Potential Index has also shown promise in predicting 1-year postrelease recidivism (Hastings, Stuewig, Drapalski, & Tangney, 2003).

Institutional misconduct. Institutional misconduct was coded from jail records obtained between the time of enrollment in the study and the release date ($n = 315$ men; $n = 123$ women). Three indices of institutional misconduct were calculated: (a) *number of incidents* recorded in the inmate's file regardless of whether they led to a formal institutional charge or finding; (b) *number of formal charges* levied against the inmate for violation of institutional rules; and (c) *number of formal physical charges* for physical violence (e.g., fighting, assault on a correctional officer, etc.) levied against the inmate. For certain analyses, variables were dichotomized to reflect any incident or charge. In the current sample, 173 (40%) inmates had at least one incident (127 men and 46 women), $\chi^2(1, N = 438) = 0.32$, *ns*; 137 (31%) inmates had formal charges against them (104 men and 33 women), $\chi^2(1, N = 438) = 1.60$, *ns*; and 32 (7%) inmates had physical charges against them, with men being significantly more likely to have a physical incident than women (29 men, 3 women), $\chi^2(1, N = 438) = 5.98$, $p < .05$. In another analysis, we considered the number of incidents of institutional misconduct ($M = 1.38$, $SD = 2.77$, for the total sample). More incidents of institutional misconduct were recorded for male inmates ($M = 1.54$, $SD = 3.02$) compared with female inmates, $M = 0.97$, $SD = 1.98$, $t(335.2) = -2.31$, $p < .05$. On average participants were incarcerated for 151 days ($SD = 108$, range = 3 to 536). Men ($M = 160$ days, $SD = 114$) were incarcerated for a significantly longer period of time than were women, $M = 124$ days, $SD = 88$, $t(285.3) = -3.59$, $p < .001$. Once length of incarceration was taken into account, there was no gender difference in number of institutional misconduct incidents $F(1, 435) = 0.73$, $p = .39$.

One-year postrelease recidivism. At 1 year postrelease, 290 participants (207 men and 83 women) were assessed, providing self-reports of arrests and self-reports of criminal behavior for which they were not arrested (violent and nonviolent). Interviewers were blind to participants' prior PCL:SV Psychopathy scores and their VRAG scores. Analyses indicated no baseline differences in gender, race, age, education, IQ, VRAG, PCL:SV Psychopathy, or the PAI scales Antisocial, Aggression, Aggression–Physical, or the Violence Potential Index between those who were followed up compared with those who were not followed up.

Four indices of self-reported recidivism were calculated: (a) *number of arrests* reflects the number of different types of crimes for which the participant was arrested (a measure of criminal versatility); (b) *number of undetected offenses* reflects the number of different types of criminal acts participants engaged in that were not detected by authorities (a measure of criminal versatility); (c) *number of arrests or undetected offenses* is a sum score of the previous two variables; and (d) *number of violent arrests or undetected offenses* reflects the number of different types of violent crimes for which the participant was arrested plus the number of different types of undetected violent criminal acts. Violent arrests and acts were defined as assault, domestic violence, robbery, murder, and kidnapping, which matches up closely with the original outcome violence variable of interest used in the development of the VRAG (all assaultive behaviors, sex offenses, armed robbery, forcible confinement, and threatening violence with a weapon). For certain analyses, variables were dichotomized to reflect any arrest or offense.

The recidivism rate at 1 year postrelease was 64%, with significantly more men ($n = 146$; 70.5%) than women ($n = 41$; 49.4%) reporting either being arrested or having committed an undetected crime, $\chi^2(1, N = 290) = 11.55$, $p < .01$. The rate of violent recidivism was 17%, with significantly more men ($n = 42$; 20%) than women ($n = 8$; 9.6%) reporting having been arrested or having committed undetected violent criminal acts, $\chi^2(1, N = 289) = 4.78$, $p < .05$.

Results

Gender Differences in Violence Risk Appraisal Guide Item and Total Score Means

We first examined whether male and female inmates differed in their scores on the VRAG. The overall VRAG total score was significantly higher for male than for female inmates (see Table 2). Of the 12 individual VRAG items, significant gender differences

Table 2
Gender Differences in Individual Violence Risk Appraisal Guide (VRAG) Items

Item	Male	Female	Inferential statistic
1. Not lived with both parents until 16 years of age	68.3%	74.5%	$\chi^2(1, N = 473) = 1.80$
2. Elementary school problems			$Z = -5.18^{**}$
None	28.7%	51.7%	
Slight	37.5%	32.4%	
Severe	33.8%	15.9%	
3. History of alcohol problems			$Z = -2.99^{**}$
None	28.7%	34.5%	
1–2	32.0%	44.1%	
3	19.5%	11.0%	
4–5	19.8%	10.3%	
4. Marital status (never married)	34.1%	15.2%	$\chi^2(1, N = 473) = 17.8^{**}$
5. Nonviolent criminal history			$Z = -0.28$
0	12.2%	13.1%	
1–2	10.1%	7.6%	
3+	77.7%	79.3%	
6. Failure on conditional release	72.3%	71.0%	$\chi^2(1, N = 473) = 0.07$
7. Age at index offense			$Z = -2.93^{**}$
39+	24.1%	40.0%	
34–38	14.9%	12.4%	
28–33	15.9%	10.3%	
27	2.4%	4.1%	
<26	42.7%	33.1%	
8. Victim injury			$Z = -2.06^{*}$
Death	0.6%	0.0%	
Hospitalized	2.1%	0.0%	
Treated/released	2.7%	1.4%	
None/slight	94.5%	98.6%	
9. No female victim	84.5%	91.7%	$\chi^2(1, N = 473) = 4.60^{*}$
10. *DSM–III–R* personality disorder	4.0%	4.1%	$\chi^2(1, N = 473) = 0.01$
11. *DSM–III–R* no schizophrenia	97.9%	98.6%	$\chi^2(1, N = 473) = 0.25$
12. Hare Psychopathy Checklist score			$Z = -5.66^{**}$
1–4	5.2%	8.3%	
5–7	6.7%	17.9%	
8–10	17.4%	25.5%	
11–16	43.9%	40.7%	
17–22	26.8%	6.9%	
23–24	0.0%	0.7%	
VRAG total score	8.5 (7.9)	4.8 (6.7)	$t(471) = -4.90^{**}$

Note. *DSM–III–R* = *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., revised). $n = 328$ (male); $n = 145$ (female). Chi-square values are presented for dichotomous items, Mann–Whitney tests are presented for ordinal items, and *t* tests are presented for continuous items.
$^{*}p < .05$. $^{**}p < .01$.

were observed on seven items. (Chi-square values are presented for dichotomous items; Mann–Whitney tests are presented for ordinal items, e.g., Items 2 and 3). Specifically, at the time of the index offense, women tended to be older (Item 7) and more likely to have been married (Item 4). They had less severe histories of elementary school problems (Item 2) and alcohol problems (Item 3), lower levels of psychopathy (Item 12), less severe injury to victims (Item 8), and the victim in their index offense was slightly less likely to be female (Item 9). We also examined gender differences on the total of the 11 VRAG items that essentially augment the Hare Psychopathy scores. Results indicated that these non–Hare-related VRAG scores for male inmates ($M = 8.05$, $SD = 6.64$) were significantly higher than for female inmates ($M = 5.61$, $SD = 5.84$), $t(471) = -3.83$, $p < .01$.

Gender Differences in Correlations Between the Violence Risk Appraisal Guide and Other Measures of Aggression

Concurrent correlations between VRAG total score and other measures of aggression or violence potential were substantially and significantly correlated for male inmates (see Table 3). Concurrent correlations between VRAG total score and other measures of aggression and violence potential were significantly correlated for female inmates as well; however, the magnitude of the correlations was smaller. Tests of independent correlations confirmed that the VRAG–PAI Antisocial, VRAG–PAI Physical Aggression, and the VRAG–PAI Violence Potential Index correlations were significantly higher for male compared with female inmates ($p < .05$).

To further investigate the source of these gender differences in the correlates of the VRAG total score, we examined the relationship of each VRAG item to the Antisocial, Aggression, Physical Aggression, and Violence Potential Scales from the PAI separately for men and women. As shown in Table 3, the VRAG items were less predictive for women compared with men. Only two VRAG items were significantly correlated with all concurrent measures of aggression for both genders: Item 2 (history of elementary school problems) and Item 12 (Hare Psychopathy). Six VRAG items showed consistently higher convergent validity with other measures of aggression for men compared with women: Item 3 (history of alcohol problems), Item 4 (never married), Item 5 (nonviolent criminal history), Item 6 (failure on conditional release), Item 7 (age at index offense), and Item 9 (no female victim). Tests of independent correlations, however, did not show many significant differences between the correlations for men compared with women on the individual items, except for Item 9 (see Table 3). The differential validity of Item 9 as a function of gender is especially notable; a high score on Item 9 (no female victim) was consistently positively correlated with measures of antisocial and aggressive behavior among men, but for women this item was consistently negatively related to other measures of antisocial behavior and aggression. Two VRAG items showed poor validity for both men and women. Item 8 (victim injury) and Item 11 (no *DSM–III–R* [*Diagnostic and Statistical Manual of Mental Disorders,* 3rd ed., revised] diagnosis of schizophrenia) were largely unrelated to concurrent measures of aggression, likely because of the minimal variance observed in these items.

Correlations and Receiver Operating Characteristics (ROCs) With Institutional Misconduct

Among male inmates, the VRAG total score was significantly correlated with number of incidents ($r = .28$, $p < .01$), number of formal charges ($r = .27$, $p < .01$), and number of formal physical charges ($r = .17$, $p < .01$), whereas among female inmates, the VRAG total score was not significantly correlated with any measure of institutional misconduct ($rs = .05$, .01, and –.04, respectively). The correlation between VRAG total score and number of incidents was significantly different between male and female

Table 3
Concurrent Correlations Between Violence Risk Appraisal Guide (VRAG) Total and Item Score Means and Personality Assessment Inventory (PAI) Scales

	PAI scale							
	Antisocial		Aggression		Aggression–Physical		Violence Potential Index	
VRAG item	Male	Female	Male	Female	Male	Female	Male	Female
1	.17**	.06	.21**	.19*	.19**	.16	.18**	.05
2	.33**	.22**	.37**	.33**	.39**	.30**	.29**	.26**
3	.29**	.11	.31**	.12	**.33****	**.12**	.34**	.20*
4	.19**	.01	.09	.02	.14**	.03	**.14****	**−.08**
5	.24**	.14	.16**	.12	.17**	.10	.25**	.11
6	.21**	.15	.17**	.16	.16**	.14	.24**	.15
7	.17**	.10	.20**	.10	.25**	.14	.11*	.01
8	.01	−.08	−.05	.01	−.03	−.02	−.05	.01
9	**.14****	**−.20***	**.07**	**−.14**	**.08**	**−.13**	.04	−.09
10	.17**	.21*	.14*	.11	.19**	.11	.14*	.16
11	−.07	−.03	−.10	−.03	−.10	.00	−.15*	−.02
12	.44**	.39**	.37**	.24**	.35**	.19*	.42**	.35**
Total	**.54****	**.37****	.50**	.39**	**.53****	**.36****	**.49****	**.32****

Note. $N = 325$–326 (male participants); $N = 143$ (female participants). Correlations that differ significantly by gender are presented in boldface.
* $p < .05$. ** $p < .01$.

inmates ($Z = -2.48$, $p < .05$) as was the correlation between VRAG total score and number of formal charges ($Z = -2.6$, $p < .05$) as well as that between VRAG total score and number of formal physical charges ($Z = -2.0$, $p < .05$).

For men, ROCs revealed moderate predictive accuracy for the VRAG total score with each of the three indices of institutional misconduct: any incident, any formal charge, and any formal physical charge (see Table 4). In contrast, ROC analyses indicated that the VRAG total score was not significantly predictive of institutional misconduct among female inmates.

To further explore these gender differences in predictive validity, we conducted correlations between each individual VRAG item and the three indices of institutional misconduct (see Supplemental Table A). Item 1 (lived with biological parents till age 16), Item 2 (elementary school problems), Item 5 (nonviolent criminal history), and Item 12 (Hare Psychopathy score) were significantly correlated with institutional misconduct for male inmates. Magnitudes ranged from small to moderate. In contrast, none of the individual VRAG items were significantly correlated with any of the three indices of institutional misconduct for female inmates. Tests of independent correlations show that men and women only significantly differed on number of incidents for Items 1 and 2, and number of formal charges for Items 1 and 12 (see Table A in the supplemental materials online).

Correlations and Receiver Operating Characteristics With 1-Year Postrelease Recidivism

The VRAG total score was significantly moderately correlated with number of arrests ($r = .25$, $p < .01$), number of undetected offenses ($r = .38$, $p < .01$), number of arrests or undetected offenses ($r = .39$, $p < .01$), and number of violent arrests or undetected offenses ($r = .37$, $p < .01$) for male inmates, whereas the VRAG total score had lower correlations with the measures of recidivism for the female inmates ($r = .21$ $p = .054$; $r = .27$ $p < .05$; $r = .27$ $p < .05$; and $r = .17$ $p = .13$, respectively). These differences between female and male inmates were not statistically significant ($Z = -.32$, $p = .75$; $Z = -.93$, $p = .35$; $Z = -1.02$, $p = .31$; and $Z = -1.64$, $p = .10$, respectively).

For men, ROCs revealed moderate predictive accuracy for the VRAG total score and all four indices of 1-year recidivism (see Table 4). In contrast, VRAG total score was only significantly related to any arrest or undetected offense for women.

Correlations between each individual VRAG item and the four indices of 1-year recidivism were conducted (see Table B in the supplemental materials). No item was consistently significantly related to all of the 1-year postrelease recidivism variables for male inmates. However, Items 2 (elementary school problems), 3 (history of alcohol problems), 5 (nonviolent criminal history), 6 (failure on conditional release), and 12 (Hare Psychopathy score) showed some consistent pattern of correlations. In general, for male inmates the magnitudes ranged from small to moderate. In contrast, only Item 12 (Hare Psychopathy score) and Item 2 (elementary school problems) were significantly correlated with any of the four indices of 1-year postrelease recidivism for female inmates. However, no correlations between female and male inmates were statistically significant from each other.

Bin Base Rates

The base rate of violent reoffending in the VRAG normative sample was 31% after 7 years. The current study base rate for violent reoffending after 1 year was 17%. The initial VRAG validation sample was divided into nine bins according to range of scores. Figure 1 shows the observed recidivism rates associated with each VRAG bin from the current male and female samples alongside the recidivism rates from the male VRAG validation sample. Despite the significantly shorter follow-up period and a limited number of subjects in Bins 1–3 and 8–9, the overall trajectory of the recidivism rates for male inmates was consistent with that of the VRAG normative sample. That is, for VRAG Bins 4–9, increasingly greater recidivism rates were observed for each bin. However, the stability of the observed recidivism rates for Bins 8–9 is limited given the small number of participants in each bin. For female inmates, scores were clustered around Bins 4–6 and were associated with a small increase in recidivism across the three bins. None of the six women in Bin 7, however, evidenced any violent recidivism. A Mann–Whitney test confirmed a

Table 4
Violence Risk Appraisal Guide Receiver Operating Characteristics With Institutional Misconduct and 1-Year Postrelease Recidivism

Variable	Area under the curve (95% *CI*)	
	Institutional misconduct during incarceration	
	Male inmates ($n = 315$)	Female inmates ($n = 123$)
Any incidents	.65** (.59–.71)	.56 (.45–.66)
Any formal charges	.70** (.64–.76)	.53 (.41–.64)
Any physical charges	.67** (.59–.75)	.43 (.04–.81)
	Recidivism at 1-year postrelease	
	Male inmates ($n = 205$–207)	Female inmates ($n = 83$)
Any arrest	.68** (.60–.75)	.62 (.49–.75)
Any undetected offense	.67** (.59–.74)	.61 (.49–.73)
Any arrest/undetected offense	.70** (.62–.78)	.66* (.54–.78)
Any violent arrest/undetected offense	.76** (.68–.85)	.66 (.47–.85)

* $p < .05$. ** $p < .01$.

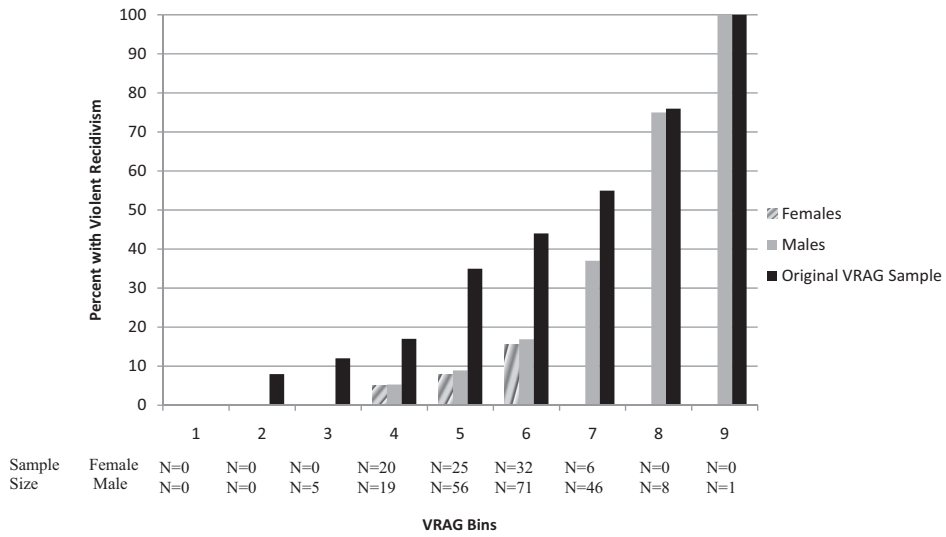

Figure 1. Observed 1-year postrelease recidivism from current sample compared with the 7-year recidivism data from original Violence Risk Appraisal Guide (VRAG) sample.

significant relationship between the bins and violent recidivism for men ($Z = -5.10$, $p < .01$) but not for women ($Z = -0.78$, $p = .43$).

The current study base rate for physical institutional misconduct was only 7%. Figure 2 shows the observed rate of physical institutional misconduct for each VRAG bin for the current sample. The low base rate of physical institutional misconduct in the sample, particularly for women where only 3 incurred physical institutional misconduct charges, rendered analysis of the VRAG at the bin level relatively uninformative.

Incremental Validity of Violence Risk Appraisal Guide Scores

To examine the incremental validity of VRAG scores we used the PCL:SV score to predict the modified version of the VRAG total score (i.e., all items summed except for the Psychopathy score item) and saved the residuals. This residual score is the independent variance of the modified VRAG once the variance associated with the PCL:SV score has been removed. Similar to Edens et al. (2006) we used this residual score in a series of ROC analyses to examine what the VRAG contributes to institutional misconduct and postrelease recidivism above and beyond the PCL:SV.

As can be seen in Table 5, the modified VRAG scores continued to predict outcomes even after controlling for psychopathy for men (although the relationship was reduced compared with the full VRAG in Table 4). For women, however, the modified VRAG scores were unrelated to outcomes. In addition to examining the dichotomous variables, we computed correlations of the modified VRAG residual with the continuous variables of institutional mis-

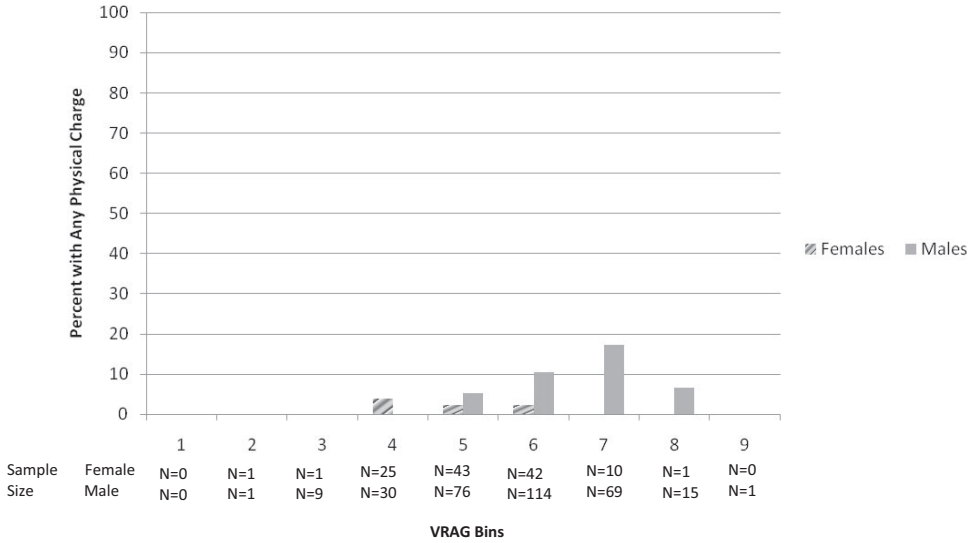

Figure 2. Observed physical institutional misconduct rates from current sample. VRAG = Violence Risk Appraisal Guide.

Table 5
Receiver Operating Characteristics of Modified Violence Risk Appraisal Guide Controlling for Hare Psychopathy Checklist: Screening Version in Order to Examine Incremental Validity

Variable	Area under the curve (95% *CI*)	
	Institutional misconduct during incarceration	
	Male inmates (n = 315)	Female inmates (n = 123)
Any incidents	.62** (.55–.68)	.50 (.40–.61)
Any formal charges	.64** (.57–.70)	.53 (.41–.64)
Any physical charges	.65** (.57–.74)	.48 (.09–.86)
	Recidivism at 1-year postrelease	
	Male inmates (n = 205–207)	Female inmates (n = 83)
Any arrest	.63** (.56–.71)	.52 (.39–.65)
Any undetected offense	.60* (.52–.68)	.50 (.38–.63)
Any Arrest/undetected offense	.60* (.52–.69)	.53 (.40–.65)
Any violent arrest/undetected offense	.71** (.62–.80)	.55 (.40–.71)

* $p < .05$. ** $p < .01$.

conduct and postrelease recidivism. For men, the residual score positively and significantly related to institutional misconduct (mean r = .14; range = .13 –.15) and to recidivism (mean r = .27; range = .21–.33) showing incremental validity. For women, the residual score was not significantly related to institutional misconduct (mean r = –.04; range = –.03 to –.04) or recidivism (mean r = .09; range = .07–.09).

Discussion

Consistent with previous research, the VRAG proved to be a moderately effective predictor of several indices of institutional misconduct and 1-year postrelease recidivism among male inmates. Most important, the VRAG—an actuarial measure designed to predict risk for violence—was a significant predictor of physical institutional charges as well as violent recidivism for men but not for women. Although not specifically validated for nonviolent offenses, the VRAG also significantly predicted general institutional misconduct as well as detected and undetected general postrelease recidivism for men, but again not for women.

The gender differences in predictive validity can be explained in part by examining the individual VRAG items and their relationship to concurrent measures of antisociality and aggression. In the current study, 10 of the 12 VRAG items were significantly correlated with concurrent measures of antisociality, aggression, and violence potential in male inmates. However, only two of the individual VRAG items (i.e., Item 2: elementary school problems; Item 12: Hare Psychopathy score) reached statistical significance in female inmates. Most of the nonsignificant individual VRAG item correlations for female inmates were in the same direction but of lesser magnitude as for male inmates; however, two items demonstrated either no relationship or an inverse relationship to concurrent measures of aggression in female inmates. Specifically, Item 9 (no female victim) was inversely related to aggression, and Item 4 (marital status) was not related to aggression in female inmates. Female-to-female violence may have a completely different meaning than male-to-female violence, thus accounting for the lack of concurrent validity for Item 9. Additionally, although marital status was a good predictor of violence and crime for men, the influence of this variable on women's behavior may be more dependent on the type of partner with whom she is in a relationship than the actual fact of being married. For example, Moffitt, Caspi, Rutter, and Silva (2001) found that as antisocial girls aged, they were more likely to become involved with antisocial men, which in turn reinforced and increased their own antisocial behavior. Research has also found that marriage serves as a protective factor against suicidality in men but not in women (Masocco et al., 2008), and being unmarried predicted lower life span in men but only to a lesser degree in women (Kaplan & Kronick, 2006).

Two VRAG items were poor predictors for both male and female inmates. Item 8 (victim injury) was largely unrelated to concurrent measures of aggression, and those who did not have a diagnosis of schizophrenia (Item 11) were often lower on aggression. Because seriously psychotic inmates were excluded from the general population, there were almost no inmates in this sample with a diagnosis of schizophrenia, which may have contributed to the contrary finding. It is possible, however, that this would be true in most general population settings, so the utility of this item may be sample specific. Nonetheless, it appears that the lack of predictive validity of VRAG scores in female inmates stems at least in part from the lack of relationship between the individual VRAG items and measures of aggression.

One limitation of the current study is that some of the outcome variables demonstrated low base rates of occurrence. However, many of the outcome base rates were low for both genders, yet the VRAG showed a medium effect size in predicting aggression among male offenders. The results of this study may also be limited to female correctional samples, as opposed to, for example, civil or forensic psychiatric samples that may vary in demographic composition and/or in base rates for some items. Finally, the reports of recidivism were all self-report, and as such there may be a bias toward false negatives in the current sample lowering validity estimates. Although official records also have many limitations of their own, it would be best to have information from multiple sources so as to guard against method bias.

The literature has largely been silent concerning the use of the VRAG with female populations, yet given a lack of actuarial risk assessment instruments normed specifically on female popula-

tions, it is not unusual for some individual forensic experts and agencies to advocate the use of the VRAG with women given the proven fallibility of unguided clinical prediction (Grove & Meehl, 1996; Meehl, 1954). The findings of this study clearly indicate that such use is inadvisable. At best, the VRAG total score provides clinicians with no more useful information about the violence risk of female inmates beyond that already accounted for by psychopathy, and at worst, the VRAG provides a distorted, less accurate assessment of women's violence risk, given its reliance on individual items that bear little predictive relevance to women's antisociality and aggression.

In short, these findings argue strongly against the use of the VRAG for assessing violence risk among female offenders. The current study demonstrates that whereas some violence risk factors are gender neutral (i.e., criminal history), clinicians cannot simply assume that many of the risk factors for violence (or risk assessment instruments) developed on male inmates will be relevant to risk assessment in female populations. One potential alternative is structured professional judgment tools such as the Historical, Clinical, Risk Management–20 and the Level of Service Inventory—Revised, which have demonstrated some reliability and validity in female samples (Folsom & Atkinson, 2007; Nicholls, Ogloff, & Douglas, 2004). Future research should also attempt to identify female-specific valid indicators of risk for violence that can augment or supplement the predictive utility of psychopathy.

References

Boccaccini, M. T., Murrie, D. C., Hawes, S. W., Simpler, A., & Johnson, J. (2010). Predicting recidivism with the Personality Assessment Inventory in a sample of sex offenders screened for civil commitment as sexually violent predators. *Psychological Assessment, 22,* 142–148. doi:10.1037/a0017818

Bonta, J., Pang, B., & Wallace-Capretta, S. (1995). Predictors of recidivism among incarcerated female offenders. *The Prison Journal, 75,* 277–294. doi:10.1177/0032855595075003002

Bureau of Justice Statistics. (2009). *Correctional populations in the United States, 2009.* Washington, DC: U.S. Department of Justice.

Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior, 36,* 567–590. doi:10.1177/0093854809333610

Coid, J., Yang, M., Ullrich, S., Zhang, T., Sizmur, S., Roberts, C., . . . Rogers, R. D. (2009). Gender differences in structured risk assessment. *Journal of Consulting and Clinical Psychology, 77,* 337–348. doi:10.1037/a0015155

Doyle, M., Dolan, M., & McGovern, J. (2002). The validity of North American risk assessment tools in predicting in-patient violent behaviour in England. *Legal and Criminological Psychology, 7,* 141–154. doi:10.1348/135532502760274756

Edens, J. F., & Ruiz, M. A. (2006). On the validity of validity scales: The importance of defensive responding in the prediction of institutional misconduct. *Psychological Assessment, 18,* 220–224. doi:10.1037/1040-3590.18.2.220

Edens, J. F., Skeem, J. L., & Douglas, K. S. (2006). Incremental validity analyses of the Violence Risk Appraisal Guide and the Psychopathy Checklist: Screening Version in a civil psychiatric sample. *Assessment, 13,* 368–374. doi:10.1177/1073191105284001

Folsom, J., & Atkinson, J. L. (2007). The generalizability of the LSI–R and the CAT to the prediction of recidivism in female offenders. *Criminal Justice and Behavior, 34,* 1044–1056. doi:10.1177/0093854807300097

Glover, A. J. J., Nicholson, D. E., Bernfeld, G. A., & Quinsey, V. L. (2002). A comparison of predictors of general and violent recidivism among high-risk federal offenders. *Criminal Justice and Behavior, 29,* 235–249. doi:10.1177/0093854802029003001

Gray, N. S., Fitzgerald, S., & Taylor, J. (2007). Predicting future reconviction in offenders with intellectual disabilities: The predictive efficacy of VRAG, PCL–SV, and the HCR-20. *Psychological Assessment, 19,* 474–479. doi:10.1037/1040-3590.19.4.474

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy, *Psychology, Public Policy, and Law, 2,* 293–323. doi:10.1037/1076-8971.2.2.293

Hare, R. D. (1991). *The Hare Psychopathy Checklist—Revised.* Toronto, Ontario, Canada: Multi-Health Systems.

Harris, G. T., Rice, M. E., & Camilleri, J. A. (2004). Applying a forensic actuarial assessment (the Violence Risk Appraisal Guide) to nonforensic patients. *Journal of Interpersonal Violence, 19,* 1063–1074. doi:10.1177/0886260504268004

Harris, G. T., Rice, M. E., & Cormier, C. A. (2002). Prospective replication of the Violence Risk Appraisal Guide in predicting violent recidivism among forensic patients. *Law and Human Behavior, 26,* 377–394. doi:10.1023/A:1016347320889

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20,* 315–335. doi:10.1177/0093854893020004001

Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D., & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment, 15,* 413–425. doi:10.1037/1040-3590.15.3.413

Hart, S. D., Cox, D. N., & Hare, R. D. (1995). *The Hare Psychopathy Checklist: Screening Version.* Toronto, Ontario, Canada: Multi-Health Systems.

Hastings, M. E., Stuewig, J., Drapalski, A., & Tangney, J. P. (2003, November). *Psychopathy, violence risk assessment, and the Personality Assessment Inventory (PAI).* Paper presented at the annual meeting of the American Society of Criminology, Denver, Colorado.

Kaplan, R. M., & Kronick, R. G. (2006). Marital status and longevity in the United States population. *Journal of Epidemiology and Community Health, 60,* 760–765. doi:10.1136/jech.2005.037606

Kröner, C., Stadtland, C., & Eidt, M. (2007). The validity of the Violence Risk Appraisal Guide (VRAG) in predicting criminal recidivism. *Criminal Behaviour and Mental Health, 17,* 89–100. doi:10.1002/cbm.644

Kroner, D. G., & Mills, J. F. (2001). The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. *Criminal Justice and Behavior, 28,* 471–489. doi:10.1177/009385480102800405

Langton, C. M., Barbaree, H. E., Seto, M. C., Peacock, E. J., Harkins, L., & Hansen, K. T. (2007). Actuarial assessment of risk for reoffense among adult sex offenders: Evaluating the predictive accuracy of the Static-2002 and five other instruments. *Criminal Justice and Behavior, 34,* 37–59.

Loucks, A., & Zamble, E. (1999). Predictors of recidivism in serious female offenders. *Corrections Today, 61,* 26–32.

Masocco, M., Pompili, M., Vichi, M., Vanacore, N., Lester, D., & Tatarelli, R. (2008). Suicide and marital status in Italy. *Psychiatric Quarterly, 79,* 275–285. doi:10.1007/s11126-008-9072-4

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis, MN: University of Minnesota Press. doi:10.1037/11281-000

Moffitt, T., Caspi, A., Rutter, M., & Silva, P. A. (2001). Sex, antisocial behavior, and mating: Mate selection and early child bearing. In T. Moffitt, A. Caspi, M. Rutter, & P. A. Silva (Eds.), *Sex differences in antisocial behavior: conduct disorder, delinquency, and violence in the*

Dunedin longitudinal study (pp. 184–197). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511490057.015

Monahan, J. (1996). Violence prediction: The past twenty and the next twenty years. *Criminal Justice and Behavior, 23,* 107–120. doi:10.1177/0093854896023001008

Morey, L. C. (1991). *Personality Assessment Inventory: Professional manual.* Odessa, FL: Psychological Assessment Resources.

Morey, L. C. (2007). *Personality Assessment Inventory: Professional manual* (2nd ed.). Odessa, FL: Psychological Assessment Resources.

Nicholls, T. L., Ogloff, J. R. P., & Douglas, K. S. (2004). Assessing violence risk among male and female civil psychiatric patients: The HCR-20, PCL:SV, and VSC. *Behavioral Sciences and the Law, 22,* 127–158. doi:10.1002/bsl.579

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk.* Washington, DC: American Psychological Association. doi:10.1037/10304-000

Salekin, R. T., Rogers, R., Ustad, K. L., & Sewell, K. W. (1998). Psychopathy and recidivism among female inmates. *Law and Human Behavior, 22,* 109–128. doi:10.1023/A:1025780806538

Skopp, N. A., Edens, J., & Ruiz, M. A. (2007). Risk factors for institutional misconduct among incarcerated women: An examination of the criterion-related validity of the Personality Assessment Inventory. *Journal of Personality Assessment, 88,* 106–117. doi:10.1207/s15327752jpa8801_14

Urbaniok, F., Noll, T., Grunewald, S., Steinbach, J., & Endrass, J. (2006). Prediction of violent and sexual offences: A replication study of the VRAG in Switzerland. *Journal of Forensic Psychiatry & Psychology, 17,* 23–31. doi:10.1080/02699200500297799

Van Voorhis, P., Wright, E. M., Salisbury, E., & Bauman, A. (2010). Women's risk factors and their contribution to existing risk/need assessment: The current status of a gender-responsive supplement. *Criminal Justice and Behavior, 37,* 261–288. doi:10.1177/0093854809357442

Walters, G. D. (2006). Risk-appraisal versus self-report in the prediction of criminal justice outcomes: A meta-analysis. *Criminal Justice and Behavior, 33,* 279–304. doi:10.1177/0093854805284409

Wonderlic. (1999). *Wonderlic Personnel Test and Scholastic Level Exam user's manual.* Libertyville, L: Wonderlic.

Received March 12, 2010
Revision received August 4, 2010
Accepted August 5, 2010 ■

Correction to Sellbom and Bagby (2010)

In the article "Detection of Overreported Psychopathology With the MMPI-2 RF Form Validity Scales," by Martin Sellbom and R. Michael Bagby (*Psychological Assessment*, Vol. 22, No. 4, pp. 757–767. doi:10.1037/a0020825), there was an error in the title. The title should have read "Detection of Overreported Psychopathology With the MMPI-2-RF Validity Scales."

DOI: 10.1037/a0022515