A MULTIPLE-MODELS APPROACH TO VIOLENCE RISK ASSESSMENT AMONG PEOPLE WITH MENTAL DISORDER

STEVEN BANKS
University of Massachusetts Medical Center
PAMELA CLARK ROBBINS
Policy Research Associates
ERIC SILVER
The Pennsylvania State University
ROUMEN VESSELINOV
HENRY J. STEADMAN
Policy Research Associates
JOHN MONAHAN
University of Virginia
EDWARD P. MULVEY
University of Pittsburgh School of Medicine
PAUL S. APPELBAUM
THOMAS GRISSO
University of Massachusetts Medical Center
LOREN H. ROTH
University of Pittsburgh School of Medicine

Actuarial models for violence risk assessment have proliferated in recent years. In this article, we describe an approach that integrates the predictions of many actuarial risk-assessment models, each of which may capture a different but important facet of the interactive relationship between the measured risk factors and violence. Using this multiple-models approach, we ultimately combined the results of five prediction models generated by the iterative classification tree (ICT) methodology developed in the MacArthur Violence Risk Assessment Study. This combination of models produced results not only superior to those of any of its constituent models, but superior to any other actuarial violence risk-assessment procedure reported in the literature to date.

CRIMINAL JUSTICE AND BEHAVIOR, Vol. 31 No. 3, June 2004 324-340
DOI: 10.1177/0093854804263635

Keywords: *multiple models; violence risk; mental disorder; assessment*

The field of violence risk assessment has recently seen a number of actuarial instruments proposed to support, or sometimes to replace, the exercise of clinical judgment. The Violence Risk Appraisal Guide (VRAG) (Harris, Rice, & Quinsey, 1993), for example, was developed from a sample of male offenders at a maximum-security hospital. A series of regression models identified 12 variables coded from institutional files for inclusion in the instrument. In a receiver operating characteristic (ROC) analysis (Swets, 1988), the area under the curve for the VRAG was .76 for any new criminal charge or return to the institution for a violent offense over a time at risk in the community that averaged 7 years (Quinsey, Harris, Rice, & Cormier, 1998). A second example is the HCR-20 (Webster, Douglas, Eaves, & Hart, 1995), a structured clinical guide derived from a review of the literature that can be scored in an actuarial manner to assess violence risk. This instrument consists of 20 ratings addressing historical, clinical, and risk-management variables. In a study of patients who were civilly committed, Douglas, Ogloff, Nicholls, and Grant (1999) found that during a follow-up of approximately 2 years after discharge, the area under the ROC curve for physical violence for the HCR-20 was also .76.

We recently described a third methodological approach to actuarial violence risk assessment, the ICT and illustrated this method by generating two different risk-assessment models from the data collected in the MacArthur Violence Risk Assessment Study (Monahan & Steadman, 1994; Steadman et al. , 1998). One model was unrestricted in terms of eligible risk factors and achieved an area under the ROC curve of .82 (Steadman et al. , 2000). Another model restricted eligible risk factors to those commonly available in hospital records or capable of being routinely assessed in clinical practice and yielded an area under the ROC curve of .80 (Monahan et al., 2000).

A concern that the relative success of the ICT at assessing violence risk might be due to overfitting the data (i.e., to capitalization on chance) has led us to estimate several different ICT models to obtain multiple risk assessments for each case. In this article, we show how multiple actuarial models can be combined to produce risk assessments that are much more accurate than any single actuarial model taken alone. In addition, we demonstrate that by scoring each individual using many different actuarial models, more participants can be categorized into groups with exceedingly high and low rates of violence.

METHOD

THE MACARTHUR VIOLENCE RISK ASSESSMENT STUDY

The methodology of the MacArthur Violence Risk Assessment Study has been described in detail elsewhere (Monahan et al. , 2001; Steadman et al. , 1998; Steadman et al. , 2000) and is only summarized here. Admissions were sampled from acute psychiatric inpatient facilities at three sites: Western Psychiatric Institute and Clinic (Pittsburgh, PA), Western Missouri Mental Health Center (Kansas City, MO), and Worcester State Hospital and the University of Massachusetts Medical Center (Worcester, MA). Selection criteria for research participants were (a) civil admissions, (b) between the ages of 18 and 40 years, (c) English speaking, (d) White or African American ethnicity (or Hispanic American in Worcester only), and (e) a chart diagnosis of schizophrenia, schizophreniform, schizoaffective, depression, dysthymia, mania, brief reactive psychosis, delusional disorder, alcohol or drug abuse or dependence, or a personality disorder.

Hospital data collection was conducted in two parts: an interview by a research interviewer to obtain data on risk factors and violence, and an interview by a research clinician (Ph.D. or MA/MSW in psychology or social work) to confirm the chart diagnosis using the DSM-IIIR Checklist and to administer several clinical instruments.

Twenty weeks after hospital discharge was chosen as the time frame for the analysis because this was the period during which the prevalence of violence by patients in the community was at its highest (Steadman et al. , 1998). Research interviewers attempted two follow-up interviews with enrolled patients in the community during this period, approximately 10 weeks apart. A collateral informant who knew of the patient's behavior in the community during the follow-up period (usually, but not always, a family member) was also interviewed on the same schedule. Arrest and rehospitalization records provided the third source of information about the patients' behavior in the community. *Violence to others* was defined to include the following: acts of battery that resulted in physical injury; sexual assaults; assaultive acts that involved the use of a weapon; or threats made with a weapon in hand.

THE ITERATIVE CLASSIFICATION TREE

The ICT method is described elsewhere (Steadman et al. , 2000). In brief, we used CHAID (Chi-Squared Automatic Interaction Detector) software (SPSS, 1993) to assess the statistical significance of the bivariate association between each of the 106 eligible risk factors measured in the hospital (see Monahan et al., 2000 for a description of these risk factors) and the dichotomous outcome measure (violence in the community) until the most statistically significant value of χ^2 was identified, with $p < .05$ a necessary condition for risk-factor selection. When a risk factor was selected, the sample was partitioned according to the values of that risk factor. This selection procedure was then repeated for each of the sample partitions, thus further partitioning the sample. The result of the partitioning process was to identify groups of cases that shared the same risk factors and that also shared the same values on the outcome measure, violence.

We then extended this recursive partitioning approach in an iterative fashion; that is, all participants not classified into groups designated as either *high risk* or *low risk* in the first iteration of CHAID were pooled together and reanalyzed in a second iteration of CHAID. This iterative process continued until it was not possible to classify any additional groups of participants as either high or low risk (with no group allowed to contain fewer than 50 cases).

We defined any group of patients with a rate of violence that was less than half the base prevalence rate of the total sample as in the low-risk category, and any group of patients whose rate of violence was greater than twice the base prevalence rate of the total sample as in the high-risk category. Because the base prevalence rate of violence during the first 20 weeks after hospital discharge for the total sample was 18.7% (Steadman et al. , 1998), this meant that the cutoff for the low-

TABLE 1: Classification Based on Two Iterative Classification Tree Models

	Clinically Feasible ICT Model							
	Low Risk		*Average Risk*		*High Risk*		*Total*	
Empirically Optimal ICT Model	N	*% Violent*	N	*% Violent*	N	*% Violent*	N	*% Violent*
Low risk	352	2.8	88	8.0	22	4.5	462	3.9
Average risk	72	6.9	75	26.7	73	28.8	220	20.9
High risk	54	13.0	94	37.2	109	64.2	257	43.6
Total	478	4.6	257	24.1	204	45.1	939	18.7

risk category was 9% violent, and the cutoff for high-risk category was 37% violent.

RESULTS

COMPARABLE OVERALL ACCURACY, BUT DIFFERENT INDIVIDUAL PREDICTIONS

Although the two illustrative ICT models we generated looked very different (i.e., different risk factors entered into the empirically optimal and the clinically feasible models), they were remarkably similar in terms of their predictive accuracy: the empirically optimal ICT model (Steadman et al. , 2000) yielded an area under the ROC curve of .82 compared to .80 for the clinically feasible ICT model (Monahan et al., 2000).

Comparable levels of predictive accuracy, however, did not imply comparable predictions for individual cases. Table 1 shows the classification of individual cases into the high- (> 37%), average- (between 9% and 37%), and low- (< 9%) risk categories by the empirically optimal and clinically feasible ICT models. As shown, of the 939 patients in the total sample, 352 (37.5%) were twice classified as low risk, 109 patients (11.6% of the total) were twice classified as high risk, and 75 patients (8.0%) were twice classified as average risk. Thus, a total of 536 patients (57.1%) received the same risk classification from both ICT models. By contrast, the numbers off the main diagonal of Table 1 represent cases whose risk classifications differed depending on which ICT model was used to make the classification. As shown, 54 patients (5.8%) were classified as low risk by the empirically optimal model, but as high risk by the clinically feasible model, and 22 patients (2.3%) were classified as low risk by the clinically feasible model, but as high risk by the empirically optimal model.

This observation (that different predictions may be obtained for the same individual from risk-assessment models that have comparable levels of predictive accuracy) is not unique to tree-based models but rather is a general property of actuarial prediction models (including main-effects prediction models; see McNiel, Lam, & Binder, 2000, on prediction by multiple clinicians). Indeed, the only circumstance under which this observation would not hold would be when the predictions made by the risk-assessment models are correlated 1.0. In this instance, however, the correlation between the predictions made by the empirically optimal ICT model and by the clinically feasible ICT model was only .52 ($p < .001$). The fact that these prediction models are comparably associated with the criterion measure, violence (as indicated by the ROC analysis), but only modestly associated with each other, suggested to us that each model taps into an important, but different, interactive process that relates to violence.

A TWO-MODEL APPROACH TO RISK ASSESSMENT

From this observation, a central set of questions emerged: What is the prevalence of violence among the cases twice classified as high risk?; and what is the prevalence of violence among the cases twice classified as low risk? In other words, what is the violence risk for cases that the empirically optimal ICT and the clinically feasible ICT classified as high risk?; and what is the violence risk for cases that the empirically optimal ICT and the clinically feasible ICT classified as low risk? Table 1 displays the rates of violence for each cell. As shown, the Low-Low group had only a 2.8% rate of violence during the 20-week follow-up and the High-High group had a 64.2% rate of violence during the same period. By contrast, the lowest and highest rates of violence we obtained with each of the ICT models separately were 3.9% and 45.1%, respectively. In addition, the area under the ROC curve for the two ICT models combined (.83) indicated a higher degree of predictive accuracy than was obtained by either ICT model operating independently.

EXPANDING THE TWO-MODEL APPROACH: MULTIPLE MODELS

If combining two models predicts violence more accurately than either model by itself predicts violence, would combining more than two models predict violence still more accurately? In expanding a two-models approach to a multiple-models approach, the primary methodological challenge lies in how to combine the results of the various models. To explore how such combination may be achieved, we constructed 10 ICT models, each of which featured a different risk factor as a starting point in building the tree.

Each of the 10 models was developed using the 106 clinically feasible risk factors described in Monahan et al. (2000), with Model 1 being the same clinically feasible ICT model described in that article. The remaining 9 ICT models were constructed using the same procedures as had been used to construct the clinically feasible ICT, with one exception: We forced a different initial variable into each of the nine trees. More specifically, the procedure involved three steps. First, we had the CHAID program list "competitor" variables to the first variable that entered into the clinically feasible ICT (which was seriousness of arrest, see Monahan et al., 2000); that is, we had the CHAID program identify those variables that would enter the ICT first if we eliminated seriousness of arrest as an eligible variable for the analysis. Second, from this list we chose nine competitors that were nonoverlapping in terms of the underlying construct being measured (i.e., we chose competitors that were not simply different indices of the same underlying variable, such as alcohol use and alcohol diagnosis). The variables chosen are listed in Table 2. Finally, we ran nine ICT analyses, each taking one of the selected variables as the initial risk factor to split the sample.

Most of the 10 ICT models required two iterations to complete, and the number of variables in each model ranged from 8 to 16 (see Table 2). Areas under the ROC curves for the 10 models varied from .73 to .81, and the percentage classified as high or low risk varied from 53.2% to 72.6%. The specific risk factors included in each model and how often each risk factor was included are listed in Table 3.[1]

TABLE 2: Characteristics of the Multiple Iterative Classification Tree Models

Model	*First Variable*	*Iterations (No.)*	*Variables (No.)*	*Classified as High or Low Risk (%)*	*Area Under ROC Curve*
1	Seriousness of arrest	3	12	72.6	.803
2	Drug abuse diagnosis	2	9	65.6	.738
3	Alcohol abuse diagnosis	2	13	60.7	.764
4	Primary diagnosis	2	8	55.3	.753
5	Anger reaction	2	11	62.8	.778
6	Schedule of imagined violence	2	10	55.8	.769
7	Child abuse	5	14	56.0	.791
8	Prior violence	3	10	74.1	.766
9	Age	2	16	53.2	.784
10	Gender	2	14	62.7	.806

Note. ROC = receiver operating characteristic.

The next step in our effort to combine the results of multiple prediction models involved dividing the risk groups produced by each of the 10 models into three categories: low-violence risk (< 9%), average-violence risk (between 9% and 37%), and high-violence risk (> 37%; see Steadman et al., 2000). We used the Ohlin-Burgess scoring method (Burgess, 1928; Ohlin, 1951) to score an individual's performance on the models: Low risk was coded as –1, average risk was coded as 0, and high risk was coded as +1.

A composite risk score was then computed for each participant by summing across the 10 models.[2] Each individual participant, therefore, had a composite risk score that could range from –10 (if the participant was in the low-risk category on all 10 models) to +10 (if the participant was in the high-risk category on all 10 models).

Actual composite risk scores ranged from –8 to 10, with a mean of 3.4, for the 176 individuals who were violent during the first 20 weeks after hospital discharge. Of all violent individuals, 75% had a score of 1 or greater, indicating that across the 10 models, they were in the high-risk category more often than they were in the low- or average-risk categories. For the 763 individuals who were not violent during the first 20 weeks after hospital discharge, composite risk scores ranged from –10 to 10, with a mean of –3.6. Although the composite risk scores covered the full range, 75% of all nonviolent individuals had a score of –1 or less, indicating that across the 10 models, they

TABLE 3: Variables Included in the Multiple Iterative Classification Tree Models

Freq./ Models	Variable	Model 1	2	3	4	5	6	7	8	9	10
10	Legal status	x	x	x	x	x	x	x	x	x	x
7	Major mental diagnosis and substance abuse	x		x	x	x	x	x			x
6	Prior arrests - frequency		x		x			x	x	x	x
6	Child abuse - seriousness		x	x		x	x	x[a]		x	
6	Diagnosis of schizophrenia, schizophreniform, schizoaffective	x	x	x				x	x	x	
6	Neurological screening - loss of consciousness	x			x		x	x	x		x
5	Age			x		x		x	x	x[a]	
5	Anger reaction	x				x[a]	x			x	x
5	Prior arrests - seriousness	x[a]	x	x			x			x	
4	Employed FT or PT	x							x	x	x
4	Schedule of imagined violence (SIV)	x	x				x[a]		x		
4	Father's drug use	x		x			x			x	
4	Lived with father until age 15		x			x		x			x
3	Alcohol abuse diagnosis			x[a]				x		x	
3	Drug abuse diagnosis		x[a]				x			x	
3	Gender		x		x						x[a]
3	Child abuse - frequency			x		x					x
3	Diagnosis of antisocial personality disorder			x		x				x	
2	Perceived coercion					x					x
2	Prior violence	x							x[a]		
2	Motor impulsiveness	x						x			
2	Parents fight with each other	x									x
2	Valid attempt to kill self			x				x			

2	Sexual abuse before age 20	x						x	
2	Marital status		x	x					
2	Threat-control override					x	x		
1	Prior hospitalization	x							
1	Thoughts of harming self								x
1	Present at admission-suicide threat							x	
1	Father's arrests								x
1	SIV - not frequent, not escalating, not while with target		x						
1	Age at first hospitalization			x					
1	Present at admission - depression				x				
1	Years of education						x		
1	Hallucinations							x	
1	Functioning score								x
1	Primary diagnosis		x[a]						
1	Present at admission - decompensate					x			
1	Present at admission - substance abuse					x			
1	Present at admission - personal problems							x	

a. First variable in the model.

TABLE 4: Violence by Combining Scores on 10 Iterative Classification Tree Models

Score	*Number of Cases*	*Violent (%)*
–10	21	0.0
–9	52	0.0
–8	86	1.2
–7	9	3.8
–6	69	0.0
–5	72	1.4
–4	66	9.1
–3	71	12.7
–2	56	12.5
–1	40	12.5
0	47	21.3
1	54	22.2
2	43	34.9
3	41	43.9
4	29	34.5
5	20	50.0
6	30	66.7
7	23	73.9
8	20	70.0
9	10	90.0
10	10	90.0

were in the low-risk category more often than they were in the high-risk category. Table 4 indicates the percentage violent for each composite risk score, ranging from .00 to .90.

As 2 models predict violence better than 1, so do 10 models predict violence better than 2 (i.e., the area under the ROC curve was .88 for 10 models compared to .83 for 2 models). However, are all 10 models necessary to achieve a high degree of predictive accuracy? To answer this question, a stepwise logistic regression was performed with violence during the first 20 weeks after discharge as the dependent measure. As shown in Table 5, only 5 of the 10 models were selected into the stepwise logistic regression equation. The overall fit was very good, $\chi^2(5, N = 939) = 300, p < .001$; $c = .878$; pseudo $R^2 = .44$ (c being the area under the ROC curve). The coefficients in these 5 models were all essentially equal, suggesting a simple summation of the scores provides predicted probabilities of violence very close to those

TABLE 5: Iterative Classification Tree Models Selected Using Stepwise Logistic Regression

Model	*First Variable*	b	*Odds Ratio*
1	Seriousness of arrest	.638	1.89*
5	Anger reaction	.530	1.70*
7	Child abuse	.705	2.02*
9	Age	.679	1.97*
10	Gender	.693	2.00*
	Constant	–1.4948	

* $p < .001$.

TABLE 6: Violence by Combining Scores on Five Iterative Classification Tree Models

Score	*Number of Cases*	*% Violent*
–5	44	0.0
–4	147	0.7
–3	152	2.0
–2	142	9.2
–1	106	5.7
0	102	25.5
1	81	27.2
2	57	52.6
3	45	60.0
4	32	71.9
5	31	80.6

produced by the weighted coefficients of the logistic regression model. The alpha for these five variables was .74.

Using a composite risk score based on the five ICT models identified in Table 5, the 176 individuals who were violent during the first 20 weeks after discharge had a mean composite score of 1.9, and 50% of all violent individuals had a score of 2 or greater. The 763 individuals who were not violent during this period had a mean composite score of –1.8, and 50% of all nonviolent individuals had a score of –2 or less. Table 6 indicates the percentage violent for each of these scores.

To increase the robustness of the multiple-model risk classifications and to produce the most parsimonious classifications possible,

TABLE 7: Clustering in Five Risk Classes

Class	*Score Range*	*Number of Cases*	*Violent %*	*95% CI (Bootstrap)*	
1	–3 or less	343	1.2	0.3	2.4
2	–1 or –2	248	7.7	4.7	11.1
3	0 or 1	183	26.2	19.5	32.4
4	2 or 3	102	55.9	46.2	65.3
5	4 or 5	63	76.2	65.4	86.2

Note. CI = confidence interval.

we used an analysis of variance to identify statistically significant differences in the percentage violent among the 11 different scores in Table 6, as well as a procedure described by Nelson (1977) to yield a monotonic violence relationship. These analyses resulted in the identification of five composite risk groups (which we will call risk "classes" to avoid confusion with the specific risk "groups" (or nodes) on an ICT and with the broad-risk "categories" created by our use of high- and low-risk cutoffs). The number of cases and percentage violent in each risk class are given in Table 7 and presented graphically in Figure 1, along with bootstrapped 95% confidence intervals. The area under the ROC curve for the final five risk classes is .88, the same figure as obtained with 10 models. To further test the value of a multiple-models approach to violence risk assessment, we asked two additional and related questions. First, is the multiple-model approach as accurate in identifying people who were repetitively violent over the 20-week follow-up period as it is in identifying people who may have been violent only once? Second, what proportion of the total violence that occurred during the 20-week follow-up was committed by people in the various risk classes?

To answer the first question, it is clear that the multiple-model approach discriminates among repetitively violent people in a significant and linear fashion. Overall, 6.9% of the participants had two or more violent incidents during the first 20 weeks after discharge. From Risk Classes 1 through 5, respectively, the percentage of participants with two or more violent acts was 0.0, 1.6, 9.7, 21.6, and 36.5 ($p < .0001$).

To answer the second question, we examined the 355 total violent acts committed by our 939 participants over the course of the 20-week

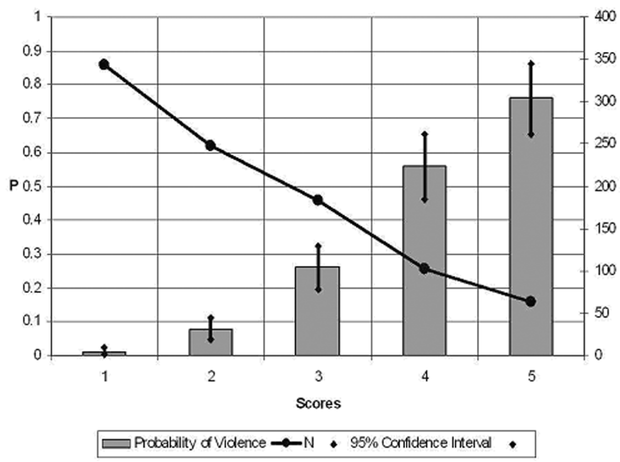

Figure 1: Number of Cases and Percentage Violent in Each Risk Class

follow-up. The results demonstrated that violence was strongly concentrated in the highest risk classes. Participants in Risk Class 1 constituted 36.5% of the sample but committed only 1.1% of the total violent acts. Likewise, participants in Risk Class 2, although making up 26.4% of the sample, committed only 7.9% of the violence. The middle-risk class (Class 3) consisted of 19.5% of the sample, and approximately the same proportion (23.7%) of the violent acts. Risk Class 4, however, although constituting only 10.9% of the sample, committed 33.8% of the total violence; and Risk Class 5, in which only 6.7% of the sample were members, accounted for 33.5% of the total violence. The two highest risk classes taken together, therefore, contained about one-sixth of the participants, and these participants committed more than two-thirds of the total number of violent acts committed by the sample.

CONCLUSION

Rather than pitting different risk-assessment models against one another and choosing the one model that appears "best," we have described an approach that integrates the predictions of many different risk-assessment models, each of which may capture a different but important facet of the interactive relationship between the measured risk factors and violence. Using this multiple-models approach, we ultimately combined the results of five prediction models generated by the ICT methodology. By combining the predictions of several risk-assessment models, the multiple-models approach minimizes the problem of data overfitting that can result when a single best prediction model is used. As important, this combination of models produced results not only superior to those of any of its constituent models, but superior to any other actuarial violence risk assessment procedure reported in the literature to date. Using only risk factors commonly available in hospital records or capable of being routinely assessed in clinical practice, we were able to place all patients into one of five risk classes for which the prevalence of violence during the first 20 weeks following discharge into the community varied between 1% and 76%, with an area under the ROC curve of .88.

The multiple-model approach to risk assessment appears to be highly accurate when compared to other approaches. However, it is also much more computationally complex than other approaches. Five ICT prediction models need to be constructed, each with between two and five iterations and each involving between 11 and 16 variables (see Table 2). It would clearly be impossible for a clinician to commit the multiple models and their scoring to memory, and using a paper-and-pencil protocol would be unwieldy in the extreme, especially because many of the risk factors appear in more than one of the models. Fortunately, however, the administration and scoring of multiple ICT models lends itself to software. In clinical use, the multiple ICTs would consist simply of a series of questions that would flow one to the next on a computer screen—through the various iterations of each of the models as necessary—depending on the answer to each prior question, much as is the case in many common diagnostic tools such as DTREE (First, Williams, & Spitzer, 1998) and the Computer-Assisted Structured Clinical Interview (SCID; First, Spitzer, Gibbon, & Williams, 1999). Under a grant from the National Institute of Mental Health, we are currently in the process of testing a prototype of such violence risk assessment software using multiple ICTs.

NOTES

1. We eliminated one variable, race, from the final models on ethical and political grounds. Race was included as an eligible variable in all 10 models but emerged from the analysis in only 3 models: 2, 7 and 8 (Table 6). In order to avoid any possible misinterpretation of our risk-assessment procedures as a form of "racial profiling," we removed the variable of race from the 3 models in which it emerged (with the next-most-statistically significant variable taking the place of race). The revised models without race differed only trivially in accuracy from the original ones that included race. For example, the area under the ROC curve for the original Model 2, which included race, was .744, while the area under the ROC curve for the revised model, with race excluded, was .738.

2. Correlations among the risk categories emerging from the 10 models—i.e., correlations among participants' low (–1), average (0), or high (+1) risk scores on the 10 models—were also computed. All models were moderately correlated with one another (from .26 to .57; all significant at $p < .001$). Internal reliability (Cronbach's alpha) was also calculated and found to be .87.

REFERENCES

Burgess, E. (1928). Factors determining success or failure on parole. In A. A. Bruce (Ed.), *The workings on the indeterminate sentence law and the parole system in Illinois* (pp. 205-248). Springfield, IL: Illinois State Board of Parole.

Douglas, K., Ogloff, J., Nicholls, T., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 Violence Risk Assessment Scheme and the Psychopathy Checklist: Screening Version. *Journal of Consulting and Clinical Psychology*, *67*, 917-930.

First, M., Spitzer, R., Gibbon, M., & Williams, J. (1999). *Computer-Assisted SCID-Clinician Version*. North Tonowanda, NY: Multi-Health Systems.

First, M., Williams, J., & Spitzer, R. (1998). *DTREE: The DSM-IV expert*. North Tonowanda, NY: Multi-Health Systems.

Harris, G., Rice, M., & Quinsey, V. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, *20*, 315.

McNiel, D., Lam, J., & Binder, R. (2000). Relevance of inter-rater agreement to violence risk assessment. *Journal of Consulting and Clinical Psychology*, *68*, 1111-1115.

Monahan, J., & Steadman, H. (Eds.). (1994). *Violence and mental disorder: Developments in risk assessment*. Chicago: University of Chicago Press.

Monahan, J., Steadman, H., Appelbaum, P., Robbins, P., Mulvey, E., Silver, E., et al. (2000). Developing a clinically useful actuarial tool for assessing violence risk. *British Journal of Psychiatry*, *176*, 312-319.

Monahan, J., Steadman, H., Silver, E., Appelbaum, P., Robbins, P., Mulvey, E., et al. (2001). *Rethinking risk assessment: The MacArthur Study of Mental Disorder and Violence* . New York: Oxford University Press.

Nelson, L. (1977). Tables for testing ordered alternatives in an analysis of variance. *Biometrika*, *64*, 333-336.

Ohlin, L. (1951). *Selection for parole*. New York: Russell Sage.

Quinsey, V., Harris, G., Rice, M., & Cormier, C. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.

SPSS. (1993). *SPSS for Windows CHAID (Release 6.0).* Chicago: SPSS.

Steadman, H., Mulvey, E., Monahan, J., Robbins, P., Appelbaum, P., Grisso, T., et al. (1998). Violence by people discharged from acute psychiatric inpatient facilities and by others in the same neighborhoods. *Archives of General Psychiatry*, *55*, 393-401.

Steadman, H., Silver, E., Monahan, J., Appelbaum, P., Robbins, P., Mulvey, E., et al. (2000). A classification tree approach to the development of actuarial violence risk assessment tools. *Law and Human Behavior*, *24*, 83-100.

Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*, 1285-1293.

Webster, C., Douglas, K., Eaves, D., & Hart, S. (1995). *HCR-20: Assessing risk for violence*. Toronto, Canada: University of Toronto, Centre of Criminology.