Law and Human Behavior, Vol. 20, No. 1, 1996

A Comparison of Actuarial Methods for Identifying Repetitively Violent Patients with Mental Illnesses

William Gardner,[1,2] Charles W. Lidz,[1] Edward P. Mulvey,[1] and Esther C. Shaw[1]

This is a progress report on the development of practical methods for the actuarial prediction of violence. The literature indicates that actuarial prediction is more accurate than clinical prediction, but in practice actuarial methods seem to be used rarely. Here we address two obstacles to the clinical use of actuarial prediction methods. First, clinicians may be averse to actuarial methods that require calculations. To remedy this, we developed a regression tree screen that presents actuarial information about violence in a series of yes/no questions. Second, using actuarial methods to identify the small minority of violent patients in a general psychiatric population may be too costly. To remedy this, we developed a method to prescreen patients for intensive evaluation using an inexpensive assessment. We evaluated regression trees and two-stage screening by comparing their accuracies against conventional actuarial methods. The results showed that actuarial predictions based on regression trees and two-stage screens were as accurate as regression-based methods in identifying repetitively violent patients. These easier-to-use methods may therefore be useful techniques for actuarial predictions.

This article is a progress report on the development of practical methods for the actuarial prediction of violence (Harris & Rice, 1994; Monahan, 1981; Mulvey & Lidz, 1984). Researchers have attempted to assist clinicians in predicting violence by generating actuarial information about associations between case factors and later violence (Monahan & Steadman, 1994). Mossman's (1994) review of clinical and actuarial prediction of violence indicated that actuarial prediction is more accurate than clinical prediction (see also Dawes, Faust, & Meehl, 1989), but in practice actuarial methods seem to be used rarely.

Here we address two obstacles to the clinical use of actuarial prediction methods. First, clinicians may be averse to actuarial predictions because the calculations required by, for example, a regression-based method may be hard to understand

[1]Law and Psychiatry Program, Department of Psychiatry, University of Pittsburgh School of Medicine.
[2]To whom correspondence should be addressed at Law and Psychiatry Research, Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine, 3811 O'Hara Street, Pittsburgh, PA 15213; e-mail: wpg@ethics.med.pitt.edu.

and hard to perform in a clinical setting. To remedy this, we explored a statistical method for generating a 'regression tree' of questions in ordinary language that produce an actuarial prediction about a case. Second, administering a battery of tests in a general psychiatric population to identify the small minority of violent patients may be too costly. To remedy this, we developed two-stage prediction methods that save costs by prescreening patients so that only those at high risk receive intensive assessments.

Regression Trees

A regression tree is a structured sequence of yes/no questions that lead to the classification of a case. For example, Fig. 1 presents a portion of a regression tree for diagnosing myocardial infarctions among patients who present in emergency rooms with chest pains (Goldman et al., 1982). The diagnosis can be made immediately if the emergency room electrocardiogram suggests infarction. If not, the tree instructs the clinician to next consider how recently the pains began, and so on. Each of the subtrees of questions attached to the "additional questions" labels at the bottom left of Fig. 1 eventually terminates in the diagnosis of the patient as having suffered a myocardial infarction, or not.

Statistical predictions requiring calculations may be infeasible in many clinical settings, while a decision procedure specified by a tree is easy to perform. A regression tree is also easy to grasp and explain because it generates a series of statements about a patient that provide reasons for the prediction. We therefore believe that clinicians will be more likely to accept regression trees than numerical formulas

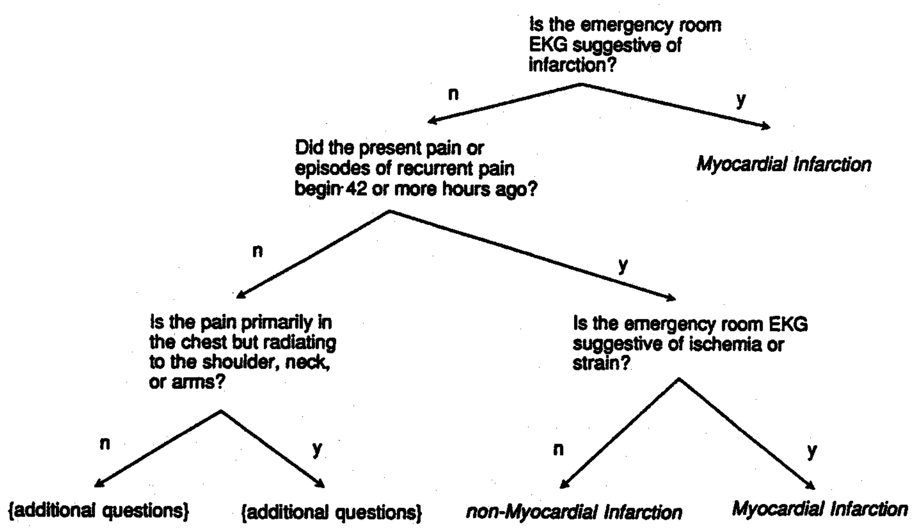

Fig. 1. Classification tree for myocardial infarction, adapted from Goldman et al. (1982).

as methods for making actuarial predictions. In this article we evaluate whether regression trees could predict violence as well as a regression-based method.

Two-Stage Screening

A practical means of predicting violence must not be too costly. Because only a few patients are frequently violent, it would be inefficient to apply an expensive screen to every psychiatric patient. We therefore devised a two-stage testing procedure. In the first stage, patients are screened using information that hospitals routinely collect. Patients selected by the first screen are then evaluated using a more intensive procedure. The question examined here is whether predictive accuracy is lost through the prescreening of patients.

Identifying Repetitively-Violent Patients

Our prediction methods were designed to identify persons with mental illnesses who are likely to be involved in high frequencies of violent incidents in the community. As in other areas of violence research (Blumstein, Cohen, Roth, & Visher, 1986), violent acts are concentrated in a small subgroup of the patient population. In the sample studied here, the modal patient (54.8% of the sample) was not violent, but the most violent 5% of the patients accounted for 45% of all incidents, and the most violent 1% accounted for 17.5%. Frequently violent patients were also more likely to have an act of serious violence, defined as an incident involving weapons, rape, attempted homicide, or an assault requiring medical treatment. Of the 25% of violent patients with the highest frequency of incidents, 41% had at least one incident of severe violence; whereas in the remaining 75% of violent patients only 24% had an incident this severe.

METHOD

Sample

The data come from a prospective longitudinal study of the accuracy of psychiatric emergency room clinicians' predictions of patients' violence (Lidz, Mulvey, & Gardner, 1993). The analysis in that study was limited to 357 pairs of patients between the ages of 14 and 65, one patient judged by clinicians to be dangerous and the other judged not to be dangerous. In addition, the pairs were matched for age, gender, race, and whether they were admitted to the hospital. The present study, however, does not require a matching design, so it can utilize additional cases who were followed in the community but could not be matched.

Patients were also selected into the sample based on the number of community interviews they completed. We attempted to interview both the patient and a collateral (someone named by the patient as likely to know what goes on in his/her life) three times over the six-month follow-up period. Cases were included if at least two community interviews with the patient had occurred, or if there were three interviews, including at least one with the patient. Twenty-eight cases, however, had to be dropped because they were hospitalized or in jail during the last four months of the study, and therefore could not have had a community incident during that period. This produced a sample of 784 cases. Of these, 13.6% had schizophrenic diagnoses, 20.0% affective disorders, 30.7% substance abuse disorders, 16.3% personality disorders, and 19.4% other diagnoses.

Adjustment for Selection Effects in the Data

To determine whether the sample was representative of the emergency room population we retrieved the gender, age, and race of the 4,713 patients seen in the emergency room during the calendar year coinciding with the primary data collection period of the study. These data showed that the sample substantially overrepresented Black males and underrepresented White females. Sixty-one percent of the sample patients were male vs. 47.1% of the ER population, and 52.6% were White vs. 63.7% of the ER (the non-White sample patients were all African-Americans). The patients in the sample were also younger (M = 28.6 years, SD = 11.1 vs. 31.6 years in the ER, SD = 12.8). These differences reflected the sampling strategy of the original study of the accuracy of clinical judgment. To obtain equal numbers of dangerous and nondangerous patients for the matching design, the recruitment procedure oversampled those emergency room patients who clinicians thought might commit violence and nondangerous individuals who had demographic characteristics similar to those patients. Thus patients likely to be involved in violence were more prevalent in the sample than they were in the emergency room population as a whole. To correct for the selection in the original design, we computed sampling weights that reflected the representation of sample patients in the ER population based on their ages, races, genders, and clinicians' judgments about their dangerousness (the method for calculating these weights is in Gardner, Mulvey, Lidz, & Shaw, 1995). These weights were used in all analyses reported here.

Covariates

Two types of covariate data were available to us. One type, the *routine* covariates, were obtained either from the patients' clinical records at the time of presentation at the emergency room, or from routine psychiatric evaluations in that setting. The routine covariates included demographics, such as age, gender, race, education, SES, living situation, and marital status. They also included clinical information available from the patients' charts, including major diagnoses, whether there was an alcohol- or substance-abuse diagnosis, whether the patient was delusional at the time of appearance at the hospital, a rating of overall stressor severity, a rating of the patients' highest level of adaptive functioning, and the history of past contacts with the hospital. Finally, the routine covariates included a count of reported violent incidents from the patient's current psychiatric evaluation and up to 5 prior evaluations, if available, as well as similar information about suicidal acts or ideation.

In addition, Lidz et al. (1993) collected data during interviews with patients in the community every two months during the six-month follow-up. Data from the first community interview were used as the *research* covariates. The research covariates included compliance with psychiatric medications, the number of drinks a patient consumed during a typical week, and the number of violent incidents during the prior 2 months. Patients were also questioned about the amount and types of drugs they had consumed during the previous two months. Based on interview questions about a spectrum of drugs, patients were classified on a 0–2 scale as nonusers, casual users, or heavy users. Casual use was defined as either less than $5 of marijuania per week, or use of only one other drug no more than once per week. Heavy drug use was defined as either (a) consumption of more than $5 per week of marijuania, (b) use of more than one drug, or (c) use of a single drug other than marijuania more than once a week.

The research covariates also included the Brief Symptom Inventory (BSI; Derogatis & Melisaratos, 1983), a 53-item self-report symptom scale. The items describe a series of problems that are rated on a 0–4 scale of degrees of distress anchored by "not at all" (0) and "extremely" (4). The subscale of the Brief Symptom Inventory that predicted violence best was Hostility (Mulvey, Gardner, Lidz, Graus, & Shaw, 1995), which included the items "Feeling easily annoyed or irritated"; "Temper outbursts that you could not control"; "Having urges to beat, injure, or harm someone"; and "Getting into frequent arguments."

The research covariates were added after exploratory data analyses convinced us that the routine covariates would not provide an adequate database to predict patients' violence. The data collected during the first community interview provided a range of detailed assessments of the patient, many of which *could* be collected during an ER assessment if they contributed substantially to the prediction of violence. Thus, for example, a patient selected for assessment on a first-stage screen based on the routine covariates might be administered the BSI and interviewed in detail about incidents of community violence during the past two months. These data would then contribute to a second stage screen. We viewed the research covariates as surrogates for the information that might have been obtained during a hypothetical intensive assessment in the ER.

Measure of Violence

Violence was ascertained from reports of incidents from the patient's psychiatric, arrest, or commitment records and from interviews with patients and collaterals (persons named by the patients as likely to know what goes on in their lives). A violent incident was one in which the patient (at least) laid hands on another person with intent to harm, or threatened the person with a weapon. Incidents did not include verbal threats, events in which the patient was the victim, parental discipline, and events in hospitals or closed treatment settings.

Because some of the actuarial screens included research covariates collected at the first community interview, we did not count violent incidents reported during that interview. The criterion to be predicted was the frequency of patient violence reported during the second and third community interviews. Thus when we report that a patient was violent during follow-up, we mean violent during the last four months of follow-up.

RESULTS

The results are organized as follows. First, a regression tree (RT) and a negative binomial regression (NBR) screen are presented and their accuracies are compared. Then a prescreen (PS) is developed from the routine data alone. Two-stage screens are formed by coupling the RT and the NBR screens with the prescreen (the PS-RT and PS-NBR screens). We then compare the accuracy of the PS-RT and PS-NBR screens against each other, and against the RT and the NBR screens alone.

The RT Screen

Figure 2 displays a tree predicting incidents per month by patients. The tree assigns a patient to one of five classes defined by a sequence of questions. For example, class 1 patients have no prior histories of violence and do not use drugs. Each class has a predicted violence rate in incidents per month (Y'_{RT}). These predictions are the rates of violence, averaged across the patients in a class, reported during the last four months of follow-up.

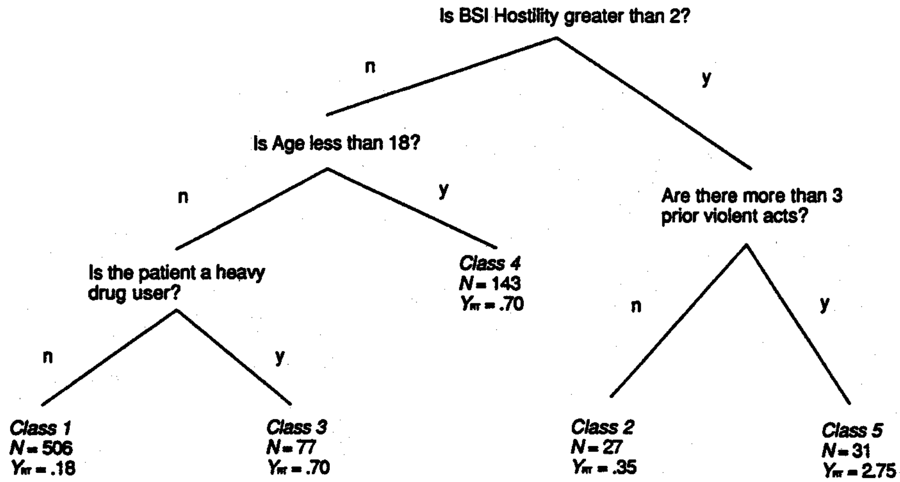

Fig. 2. RT screen for prediction of community violence by persons with mental illness.

The tree in Figure 2 was generated as follows. In the first step, the Classification and Regression Trees algorithm (CART; see Breiman, Friedman, Olshen, & Stone, 1984; Clark & Pregibon, 1992; Efron & Tibshirani, 1991) was used to generate a tree on an actuarial basis. The CART algorithm uses predictor variables to successively split the sample into subgroups that are as homogeneous as possible on the counts of violent incidents. Beginning with the complete sample (the top node in Fig. 2), the algorithm examined all possible cutpoints on each of the research and routine covariates to identify that split of the sample that created frequent- and infrequent-violence subgroups having minimum within-group variation in violence rates. The first split was whether the patient endorsed more than two of the Hostility items on the BSI (only 58 patients did). Having split the sample on Hostility, the algorithm recursed and attempted to find the optimal split of each of the Hostility subgroups. The algorithm continued to split subgroups until all groups either had no variance on the incident count or had five or fewer members. This preliminary tree was far more finely split than the one in Fig. 2. It was also excessively complex, in that many of the finely split tree's predictions would likely be false in fresh data because of capitalization on chance in the choice of splits in the preliminary tree. In a second step, therefore, the CART algorithm recollapsed many of the splits into a smaller but more parsimonious tree. The size of that 'pruned' tree was determined by computing a cross-validated estimate of the rates prediction errors likely to be associated with recollapsed trees of increasingly smaller sizes, and selecting the tree with the minimum estimated rate of prediction errors. This cross-validation procedure produced a few subgroups that continued to concern us as possibly reflecting capitalization on chance (for example, 14 year olds were split from older teenagers as more likely to be violent). We therefore simplified the tree still further by collapsing splits that produced subgroups with 25 or fewer cases, producing the result in Fig. 2.

The tree in Fig. 2 does not include several variables that predict violence on a univariate basis, such as alcohol consumption. These variables are absent from the tree because they could not be used to split any of the groups at the bottom of the tree in a way that reduced the cross-validated error rates. This is the same phenomenon one observes in multiple regression: variables that predict on a bivariate basis often do not enter a regression equation when other correlated predictors are in the equation.

The tree produced classes of patients with different levels of predicted violence, but which classes of patients should be labeled high risk? We can divide the classes into low- or high-risk groups by identifying a cutpoint or threshold rate of predicted violence, such that all those patients who are members of classes with predicted violence rates above the threshold would be labeled high risk. The RT predicts only four distinct violence rates, and thus there are only three cutpoints that matter. Let cutpoint A equal .2 incidents/month. If cutpoint A were used to delineate high risk, then class 1 patients would be low risk and classes 2–5 (37.9% of the patient population) would be high risk. If cutpoint B = .4 incidents/month were used, then classes 1 and 2 would be low risk and classes 3, 4, and 5 would be high risk (33.8% of the population). If cutpoint C = .8 incidents/month were used, only class 5 would be identified as high risk (a little over 3% of the population). Cutpoints A and B would identify too many patients as high risk, but the small size and extreme predicted violence (more than 30 incidents/year) of the group identified by cutpoint C reflects the degree to which violence is highly concentrated in a small subset of patients.

The Negative Binomial Regression (NBR) Screen

An actuarial predictor based on a regression model was developed to provide a benchmark for the accuracy of the RT. Because counted data do not fit the assumptions of ordinary linear regression, we developed a screen (Table I) using negative binomial regression (Gardner, Mulvey, & Shaw, 1995; McCullagh & Nelder, 1989), a method specifically designed for counted data. The variables included in the screen were two routine covariates—age and the number of prior incidents when seen in the emergency room (PRIOR INCIDENTS)—and three research covariates—BSI hostility, drug use during the first two months in the community, and incidents during the first two months in the community (INCIDENTS_1).

Comparison of the Accuracies of the RT and NBR Screens

Rates of Violence Among Patients Identified as High Risk

The first criterion of predictive accuracy was the rate of violence (in incidents/month) among patients identified as being at high risk. A bootstrap analysis (Efron, 1986; Gardner et al., 1995) was used to calculate nonparametric estimates of the rates of violence among patients identified as high risk by the RT, and the sampling variation of those rates. The structures labeled A, B, and C in Fig. 3 represent the sampling distributions of violence rates in the RT high-risk groups at each cutpoint. The dots near the midpoints of the vertical lines are the average (across 200 bootstrapped samples) violence rates. The coordinates of these dots on the horizontal axis are the average sizes of the high-risk groups identified by the RT, expressed as a percentage of the population. The lower bars and upper bars identify the 5% and 95% quantiles of the bootstrapped cumulative distribution of the violence rates. Together they bound a 90% bootstrap confidence interval. Thus

Table I. The NBR Screen

Covariate	Value	*SE*	*t*	*p*
AGE	-.044	.0094	-4.71	.001
log(1 + PRIOR INIDENTS)	.32	.11	3.08	.001
INCIDENTS_1	.11	.023	4.75	.001
HOSTILITY	.31	.11	2.90	.004
DRUG USE	.51	.14	3.66	.001

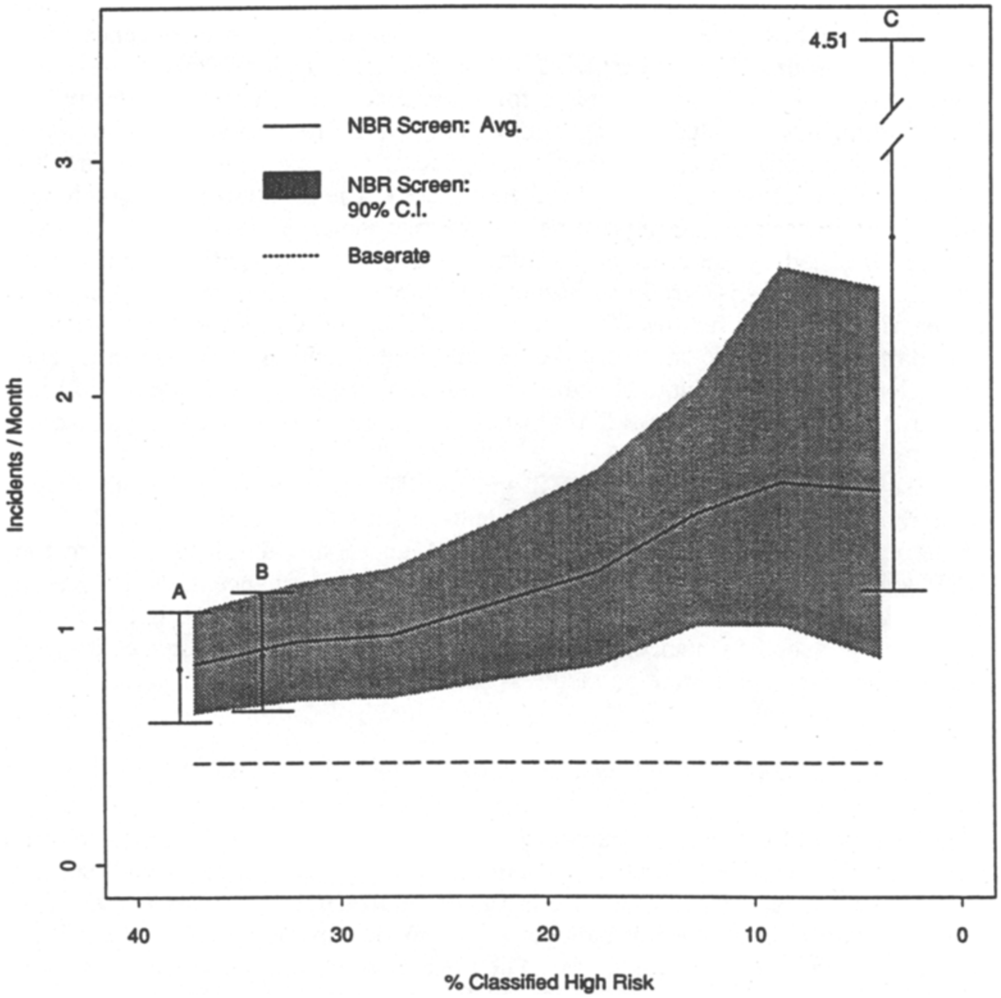

Fig. 3. Bootstrap evaluation of the RT and NBR screens: Incidents per month by patients classified as high risk by each method. The broken line in the sampling distribution for RT cutpoint C indicates that the upper bound of the distribution extends beyong the top of the graph.

at cutpoint C the RT identifies a smaller but far more violent subgroup of patients than it does at cutpoints A or B.[3]

[3]The accuracy statistics reported in this article are not cross-validated. Classical cross-validation of a prediction model requires estimating the model on a subset of the data and validating it on the rest. As is often noted, however, this wastes information that ought to be used estimating the model. A previous bootstrap cross-validation of a logistic regression prediction using the same cases and same variables as the NBR screen (Gardner et al., 1995) indicated that the optimism in the estimated sensitivity and specificity statistics attributable to capitalization on chance amounted to 1.5% of thoses statistics' values, on average. This small degree of capitalization on chance is consistent with theoretical results of Efron (1986), and is attributable to our conservative strategy of including few covariates in the predictive model relative to the large sample size.

Figure 3 also presents the estimated sampling distribution of violence rates in high-risk groups identified by the NBR screen. NBR high-risk groups were delineated by choosing seven cutpoints for NBR predicted scores. These cutpoints identified high-risk groups ranging from about 35% of the emergency room population to about 5%. The NBR was then applied at each cutpoint to the same 200 bootstrap data sets that had been used with the RT. The solid line running through the shaded area in Fig. 3 represents the average violence rates in the high-risk groups identified by the NBR at this range of cutpoints, graphed as a function of the size of the high-risk group identified at a cutpoint. This line may be compared to the violence rates for the RT at cutpoints A, B, and C. The lower and upper boundaries of the shaded area are the 5% and 95% quantiles of the violence rates in the NBR high-risk groups. The horizontal dashed line is the base rate of violence among patients: the violence rate that would be obtained by identifying "high-risk" patients purely by chance.

Thus Fig. 3 shows that the NBR and the RT screen performed much better than chance in identifying frequently violent patients. The sampling distributions of the violence rates for the RT at cutpoints A and B are almost identical to the NBR screen at comparable high-risk group sizes. The performance of the RT screen at cutpoint C is better, on average, than that of the NBR screen, but the high sampling variance of violence rates in these small high-risk groups makes it questionable whether the advantage of the RT is statistically reliable.

The Sensitivity and Specificity of the RT and NBR

The second criterion concerned the accuracy of predicting the simple occurrence of violence (Hart, Webster, & Menzies, 1993). The ideal predictive method would have low rates of false positive errors or, equivalently, high specificity, which is the percentage of not-violent patients whose predicted violence (Y') was less than the cutpoint that delineated high risk. The ideal method would also have low rates of false negative errors or, equivalently, high sensitivity, which is the percentage of violent patients whose predicted violence was greater than the cutpoint that delineated high risk. The sensitivity and specificity of the predictive methods therefore vary depending on the choice of cutpoint. Moreover, there is a tradeoff between sensitivity and specificity. High sensitivity (low rates of false negative errors) can only be achieved by choosing cutpoints for Y' that label a large group of patients as being at high risk. This produces a substantial false positive error rate, that is, low specificity. Conversely, high specificity can be achieved only with a high cutpoint, resulting in a small high-risk group and a substantial false negative error rate. Thus we will compare the actuarial methods by examining the tradeoffs that each offers between sensitivity and specificity across their ranges of cutpoints.

To evaluate the overall quality of the NBR screen we calculated the area under its ROC curve (the plot of the sensitivity of a test against its specificity for the range of the test's possible cutpoints; see Gardner et al., 1995; Rice & Harris, 1995). The ROC curve for the NBR enclosed 70.8% of the area in Fig. 3 (*SD* = 2.95%). Based on analyses in Gardner et al. (1995), however, this estimate of the area under the curve may be 2% or so too large as a result of capitalization on chance. Mossman (1994) reported an estimate of 71.3% for the areas under the ROC curve for previous actuarial predictors of patient violence, suggesting that the performance of the NBR screen was similar to that of previously reported actuarial methods. Because the RT makes predictions at only one point of interest (C, in Fig. 3) a meaningful ROC for the RT could not be plotted. However, one can compare the RT's sensitivity (7.7%, SD = 2.1%) and specificity (99.2%, SD = .5%) at cutpoint C with the most similar cutpoint examined for the NBR screen (9.3%, SD = 2.3%, and 99.1%, SD = .5%). These rates differ by no more than a standard deviation so we concluded that they were similar.

Two-Stage Screening

Prescreening Patients. An obstacle to applying either the NBR or RT screen in routine screening of patients is that each would require additional testing (the administration of the BSI) or interviewing (about recent violence and drug use). Because most patients are not dangerous, it should be possible to prescreen patients using covariates that are routinely collected in the emergency room so as to avoid testing a substantial portion of those patients who are unlikely to be violent.

The Prescreen (PS). The PS is a negative binomial regression model based on the routine covariates. The overall significance of the model was $\chi^2(3) = 104.9$, $p < .001$. The PS includes just three predictors. Being young (the covariate is log[AGE – 13], $B = -.40$, $t(780) = -3.98$, $p < .001$) predicts violence, as does having many prior incidents (log[1 + PRIOR_INCIDENTS], $B = .42$, $t(780) = 3.70$, $p < .001$). A variable THOUGHT-DISORDER was coded 1 for patients who had either a schizophrenic diagnosis or reported delusions at the time of appearance at the hospital, and zero if they had neither ($B = -.59$, $t(780) = -1.95$, $p < .06$). Thought-disordered patients were *less violent* (17.6% had at least 1 incident during the last 4 months in the community) than patients who had neither schizophrenic diagnoses nor reports of delusions (41.7% with at least one incident; the difference is statistically significant with $\chi^2(1, N = 784) = 1{,}759$, $p < 10^{-5}$). This covariate becomes only marginally significant when both age and prior incidents are in the equation, because thought-disordered patients are older (M = 36.6 and SD = 11.7 years, versus M = 27.6 and SD = 12.6 years for those without thought disorders) and have fewer prior incidents (M = 1.94 and SD = 3.34 incidents, versus M = 2.83 and SD = 5.05 incidents for those without thought disorders).[4]

[4]This pattern is interesting, because other researchers (Link, Andrews, & Cullen, 1993; Swanson, Holzer, Ganju, & Jono, 1990) have found that the presence of psychotic symptoms was a risk factor. The critical difference, we believe, is that Link's and Swanson's findings are based on epidemiological samples of patients and nonpatients whereas our data include only emergency room patients. Harris and Rice (1994) also found that schizophrenic patients were less violent than others in a sample of hospitalized criminal offenders.

Table II. Comparisons Among One- and Two-Stage Screens

		Screens			
		One-stage		Two-stage	
Criterion	Statistic	Tree	NBR	Tree	NBR
Risk group size	M	3.1%	3.8%	2.9%	2.7%
	SD	.8%	.9%	.8%	.7%
	High 5% C.I.	4.3%	5.2%	4.2%	4.0%
	Low 5% C.I.	1.8%	2.5%	1.5%	1.7%
Incidents/month	M	10.7	6.4	10.1	7.0
	SD	4.2	2.0	4.3	2.3
	High 5% C.I.	18.0	9.8	18.7	11.4
	Low 5% C.I.	4.6	3.5	4.5	3.6
Sensitivity	M	7.7%	9.3%	6.9%	6.9%
	SD	2.1%	2.3%	2.1%	1.9%
	High 5% C.I.	10.9%	12.7%	10.3%	10.2%
	Low 5% C.I.	4.2%	5.6%	3.7%	3.8%
Specificity	M	99.2%	99.1%	99.3%	99.5%
	SD	.5%	.5%	.5%	.4%
	High 5% C.I.	100%	99.8%	100%	100%
	Low 5% C.I.	98.4%	98.2%	98.4%	98.5%

Comparison of the Accuracies of the Two-Stage Screens, the RT, and the NBR

The accuracies of both two-stage screens—the PS preceding either the RT or NBR—were estimated by applying both stages to bootstrap data sets drawn from our sample. Patients who were predicted to be repetitively violent at both stages of the screen were identified as high risk. Violence rate, sensitivity, and specificity statistics were calculated as before. For simplicity, we report the comparisons of one- and two-stage trees and NBR screens (Table II) that have been calibrated to select about the same percentage of the populations as was chosen by the one-stage RT at cutpoint C. In each case, the accuracy of the PS-NBR screen was similar to that of the NBR screen alone; and similarly, the accuracy of the PS-RT screen was similar to that of the RT screen alone.

DISCUSSION

This study contributes two findings to the development of actuarial methods for the prediction of violence among mental patients. The screens developed did not demonstrate greater accuracy than prior actuarial screens (Mossman, 1994), and we do not recommend these procedures for routine clinical use. (We have reported elsewhere, however, that these screens are more accurate than unaided clinical judgments; see Gardner, Lidz, Mulvey, & Shaw, in press.) We were, however, successful in showing that actuarial methods could be implemented in ways that could lead to wider use of actuarial techniques. First, predictions based on regression trees developed using the CART algorithm were as accurate as predictions computed from regression equations. The similar accuracy of the tree and numerical regression methods is consistent with the results of a previous comparison of statistical prediction methods (Hadorn, Draper, Rogers, Keeler, & Brook, 1992). The second finding was that screening can be implemented using a two-stage strategy that is less costly than applying an intensive screen to each patients. A prescreen using only routinely-collected psychiatric data can be used to cheaply winnow out those patients who are unlikely to be violent. Then a more accurate and more expensive screen can be used with the remaining patients to identify persons at high risk for repetitive violence. Up to 90% of patients could be prescreened out with no detectable loss of accuracy.

In summary, this article shows that actuarial predictions can be generated by using a simple list of yes/no questions, and that special batteries of tests can be avoided for all but a small minority of high-risk individuals. Both of these goals can be accomplished without a significant tradeoff in the accuracy of prediction. We believe that these techniques have the potential of making actuarial prediction of violence more understandable, acceptable, and efficient in routine clinical use.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institute of Mental Health and the United States Secret Service (NIMH grant No. MH 44030-06). Support was also given to W. Gardner from the John D. and Catherine T. MacArthur Foundation Network on Mental Health and the Law. Portions of this article were first presented at the Network's Meeting on Risk Reduction in February, 1994. We thank David Banks for statistical advice and William Venables for programming the negative binomial regression module. The authors thank all of the staff, who collected and coded these data with exceptional diligence.

REFERENCES

Blumstein, A., Cohen, J., Roth, J. A., & Visher, C. A. (1986). *Criminal careers and "career criminals"*. Washington, DC: National Academy Press.

Breiman, L., Friedman J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Pacific Grove, CA: Wadsworth & Brooks/Cole.

Clark, L. A., & Pregibon, D. (1992). Tree-based models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical models in S* (pp. 377–420). Pacific Grove, CA: Wadsworth & Brooks/Cole.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1674–1688.

Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An introductory report. *Psychological Medicine, 13,* 595–605.

Efron, B. (1986). How biased is the apparent error rate of a prediction rule? *Journal of the American Statistical Association, 81,* 461–470.

Efron, B., & Tibshirani, R. (1991). Statistical data analysis in the computer age. *Science, 253,* 390–395.

Gardner, W., Lidz, C. W., Mulvey, E. P., & Shaw, E. C. (1995). Clinical versus actuarial predictions of the seriousness of violence by patients with mental illnesses. *Journal of Consulting and Clinical Psychology.*

Gardner, W., Mulvey, E. P., Lidz, C. W., & Shaw, E. C. (1995). Debiasing receiver operating characteristic curves using bootstrapping, with an application to the prediction of violence. *Submitted for publication.*

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin, 118,* 392–404.

Goldman, L., Weinberg, M., Weisberg, M., Olshen, R., Cook, E., Sargent, R., Lamas, G., Dennis, C., Wilson, C., Deckelbaum, L., Fineberg, H., Stiratelli, R., & the Yale-New Haven Brigham and Women's Medical House Staffs (1982). A computer-derived protocol to aid in the diagnosis of emergency room patients with acute chest pain. *New England Journal of Medicine, 307,* 588-596.

Hadorn, D. C., Draper, D., Rogers, W. H., Keeler, E. B., & Brook, R. H. (1992). Cross-validation performance of mortality prediction models. *Statistics in Medicine, 11,* 475–489.

Harris, G. T., & Rice, M. E. (1994). The violent patient. In R. T. Ammerman & M. Hersen (Eds.), *Handbook of prescriptive treatments for adults.* New York: Plenum.

Hart, S. D., Webster, C. D., & Menzies, R. J. (1993). A note on portraying the accuracy of violence predictions. *Law and Human Behavior, 17,* 695–700.

Lidz, C. W., Mulvey, E. P., & Gardner, W. (1993). The accuracy of predictions of violence to others. *Journal of the American Medical Association, 269,* 1007–1011.

Link, B. G., Andrews, H., & Cullen, F. T. (1993). Reconsidering the violent and illegal behavior of mental patients. *American Sociological Review, 57,* 1229–1236.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

Monahan, J. (1981). *The clinical prediction of violent behavior.* Washington, DC: United States Government Printing Office.

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

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62,* 783–792.

Mulvey, E. P., Gardner, W., Lidz, C. W., Graus, J., & Shaw, E. C. (1995). Symptomatology and violence among mental patients. *Submitted for publication.*

Mulvey, E. P., & Lidz, C. W. (1984). Clinical considerations in the prediction of dangerousness in mental patients. *Clinical Psychology Review, 4,* 379–401.

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology.*

Swanson, J. W., Holzer, C. E., Ganju, V. K., & Jono, R. T. (1990). Violence and psychiatric disorder in the community: Evidence from the Epidemiological Catchment Area surveys. *Hospital and Community Psychiatry, 41,* 761–770.