

# Criminal Justice and Behavior

<http://cjb.sagepub.com>

---

## **Classifying Juvenile Offenders According to Risk of Recidivism: Predictive Validity, Race/Ethnicity, and Gender**

Craig S. Schwalbe, Mark W. Fraser, Steven H. Day and Valerie Cooley

*Criminal Justice and Behavior* 2006; 33; 305

DOI: 10.1177/0093854806286451

The online version of this article can be found at:  
<http://cjb.sagepub.com/cgi/content/abstract/33/3/305>

---

Published by:



<http://www.sagepublications.com>

On behalf of:

International Association for Correctional and Forensic Psychology

**Additional services and information for *Criminal Justice and Behavior* can be found at:**

**Email Alerts:** <http://cjb.sagepub.com/cgi/alerts>

**Subscriptions:** <http://cjb.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations** <http://cjb.sagepub.com/cgi/content/refs/33/3/305>

# CLASSIFYING JUVENILE OFFENDERS ACCORDING TO RISK OF RECIDIVISM

## Predictive Validity, Race/Ethnicity, and Gender

CRAIG S. SCHWALBE

*Columbia University*

MARK W. FRASER

STEVEN H. DAY

VALERIE COOLEY

*University of North Carolina at Chapel Hill*

---

---

Structured risk assessment instruments are increasingly used in juvenile justice systems to support judicial decision making. They help juvenile justice authorities identify youths with a higher likelihood of repeat delinquency and reduce discretion associated with disposition decision making. To be effective, these instruments should be accurate across diverse populations. This study describes the predictive validity of the North Carolina Assessment of Risk (NCAR) in a sample of 9,534 adjudicated juveniles in North Carolina. Results show the predictive validity of the NCAR to differ by gender and race/ethnicity. Closer inspection reveals that risk factors for recidivism differed according to demographic group and that brief risk assessment instruments such as the NCAR leave other risk factors unmeasured. The results support the utility of risk assessment for juvenile justice decision making and suggest strategies to improve the validity of risk assessment for all offender groups.

**Keywords:** risk assessment; juvenile justice; juvenile court; delinquency; structured decision making

---

---

**AUTHORS' NOTE:** *The authors would like to thank Catherine Dudley, Donn Hargrove, the staff of the North Carolina Department of Juvenile Justice and Delinquency Prevention, and the Governor's Crime Commission for their assistance in completing this project. Correspondence concerning this article should be addressed to Craig S. Schwalbe, Columbia University School of Social Work, 1255 Amsterdam Avenue, New York, NY 10027; e-mail: css2109@columbia.edu.*

CRIMINAL JUSTICE AND BEHAVIOR, Vol. 33 No. 3, June 2006 305-324

DOI: 10.1177/0093854806286451

© 2006 American Association for Correctional and Forensic Psychology

Structured risk assessment instruments have been adopted widely by juvenile justice systems across the United States and Canada (Hoge, 2002). These instruments are intended to reduce, in part, individual discretionary biases that are often associated with court decision making and case dispositions. Studies have shown that decisions regarding juveniles with comparable instant offenses and prior records often vary by factors such as race/ethnicity and gender (Chesney-Lind & Shelden, 1998; Pope & Feyerherm, 1995). Risk instruments are designed to reduce racial, ethnic, and gender disparities and biases by increasing the consistency of assessment through a structured process. In practice, the result should be an increase in the reliability or consistency of case decisions made by juvenile justice officials responsible for intake, disposition, release, and other decisions (Baird, Wagner, Healy, & Johnson, 1999; Schwalbe, Fraser, Day, & Arnold, 2004). Thus, coupled with needs assessment, sentencing guidelines, and other reforms, risk assessment is an important element of a larger strategy to reduce both racial and gender disparities in the treatment of offenders by the juvenile justice system.

Although the salutatory effects of increasing consistency are well noted, the goal of promoting racial, ethnic, and gender equity (i.e., reducing disparities based on race/ethnicity and gender) may remain unfulfilled if the predictive validity of risk assessment instruments differ by gender or race/ethnicity. Substantial differences in predictive validity across demographic groups may inadvertently introduce systematic biases rather than neutralize them. Indeed, risk assessment instruments with differential validity by race/ethnicity or gender could do more harm than good if the authority of standardized measurement seemed to verify existing biases. To estimate the extent of this problem, this study investigated the predictive validity of one brief risk assessment instrument, the North Carolina Assessment of Risk (NCAR).

#### RISK ASSESSMENT IN JUVENILE JUSTICE

The purpose of risk assessment in juvenile justice is to predict future offending. Risk assessment instruments do this by measuring diverse risk factors for recidivism. Two extensive empirical reviews inform on this effort. Lipsey and Derzon (1998) reviewed 34 longitudinal studies

of the correlates of serious criminal or violent behavior, defined as any index offense, whereas Cottle, Lee, and Heilbrun (2001) focused on 22 studies of the risk factors associated with rearrest, readjudication, probation violation, and recommitment. Both studies found risk factors with strong effects in the domains of offense history, psychological factors, family factors, and peer factors. Most risk assessment instruments incorporate risk factors from two or more of these categories.

In most risk assessment instruments, scores from risk factors are added together to derive a cumulative risk score, which is reclassified into three or more risk classes (e.g., low risk, medium risk, high risk). These classifications correspond to an array of graduated sanctions and court interventions designed to prevent recidivism (Howell, 1995, 2003). The NCAR is an example in which a cumulative risk score based on nine risk factors is classified into one of three risk groups: low, medium, or high. Instruments such as the NCAR have high levels of predictive validity when juveniles are classified into risk classes with widely varying rates of recidivism.

Risk assessment instruments should correctly classify offenders across diverse demographic groups. For example, a medium-risk classification should convey a similar meaning with respect to the probability of recidivism for males, females, African American juveniles, and White juveniles. Under ideal conditions, group differences in rates of reoffending should be accounted for by differences in the proportions of juveniles classified into lower and higher risk categories. In statistical modeling terms, we would expect that the effects of race/ethnicity or gender on recidivism would be mediated by cumulative risk level.

Four published studies have directly compared the predictive validity of risk assessments across race/ethnicity and gender. Validation studies of two comprehensive instruments, the Youth Level of Service/Case Management Inventory (42 items;  $N = 250$ ; 49% Native Canadian, 34% female) and the Young Offender Level of Service Inventory (76 items;  $N = 164$ ; 50% female) found no differences across demographic groups (Ilacqua, Coulson, Lombardo, & Nutbrown, 1999; Jung & Rawana, 1999). In contrast, validation studies of two brief instruments, the NCAR (9 items;  $N = 464$ ; 50% African American, 25% female) and a local risk assessment instrument used in Orange County, California (10 items;  $N = 159$ ; 33% female),

found that predictive validity was higher for males than for females (Schwalbe et al., 2004; Sharkey, Furlong, Jimerson, & O'Brien, 2003). In addition, Schwalbe et al. (2004) found that predictive validity of the NCAR was higher for White juveniles than for African American juveniles. Although the literature is meager and conclusions necessarily tentative, these studies hint that more comprehensive measures of risk have more equivalent levels of predictive validity across gender and race/ethnicity than brief instruments.

Several explanations for this trend are possible, including (a) omitted variable bias, (b) dimensional identity, and (c) sampling bias. Each of these represents a threat to the capacity of brief risk assessment instruments to correctly and equivalently classify risk of recidivism for juveniles from diverse groups.

Omitted variable bias occurs when excluded risk factors are disproportionately distributed across populations. When omitted from risk assessment instruments, these variables may retain their influence and show up in gender and race/ethnicity parameter estimates. Statistically significant parameter estimates for demographic variables are problematic because they show that risk assessment instruments have not accounted for all of the risk-related variance that affects juveniles. That is, they show that risk assessment classifications are variably accurate across race/ethnicity and gender. Examples of risk factors often omitted from risk assessment instruments, especially from brief instruments, include neurological risks (such as hyperactivity/inattention problems) that affect boys in greater numbers than girls (Lahey et al., 1999) and neighborhood or other contextual risks that affect minority youths in greater proportion than White youths (Loeber & Farrington, 1999).

Risk assessment instruments may also succumb to the dimensional identity problem. Dimensional identity is a concept developed in the person-centered research paradigm in developmental psychology and may be useful for understanding risk assessment with juvenile offenders (Von Eye & Bergman, 2003). Essentially, the property of dimensional identity is said to exist when a measure, such as a risk assessment instrument, has the same relationship to recidivism for all subpopulations within a sample. When the empirical relationship between a risk assessment instrument and recidivism is the same for males, females, White offenders, and minority offenders,

then dimensional identity is established and the generalizability of the risk assessment instrument can be asserted. When the empirical relationship between a risk assessment instrument and recidivism differs according to subgroups, dimensional identity does not exist and the generalizability of risk assessment across demographic groups cannot be asserted. The consequence of the dimensional identity problem is considerable for juvenile justice researchers and practitioners. If a risk assessment instrument fails to possess dimensional identity, then the predictive validity of risk assessment will be greater for some groups and less for others.

A third potential threat to risk assessment predictive validity is sampling bias caused by juvenile justice decision-making practices. Differential treatment by race/ethnicity and gender affects the ability of the juvenile justice system to detect delinquency and the opportunity of juveniles to engage in delinquent activities. Research has documented, for instance, that surveillance by law enforcement is greater for African American youths compared to White youths (in at least some jurisdictions) and that, controlling for type of delinquent activity, girls are less likely to experience arrest than males (Brownfield, Sorenson, & Thompson, 2001; Chesney-Lind & Shelden, 1998; Leiber & Mack, 2003; Leiber & Stairs, 1999; MacDonald & Chesney-Lind, 2001). Once juveniles penetrate the juvenile justice system, African American youths are more often confined to secure detention facilities, and females are more likely to face harsher sanctions for less serious delinquency (MacDonald & Chesney-Lind, 2001; Pope & Feyerherm, 1995). Unequal surveillance biases risk assessment by introducing systematic error into the most frequently used criterion variable: referral to juvenile court. Because the delinquency of males and of African American youths is more likely to be officially detected, female and European American youth recidivism rates will be artificially suppressed, potentially changing the relationship between measured risk and recidivism.

Any of these explanations could limit the predictive validity of a single risk assessment instrument for a diverse sample. This is problematic because, if predictive validity varies across key subgroups, then the contribution of structured risk assessment to the laudable goal of decision-making equity will be attenuated. In particular, the match between juveniles and a graduated array of sanctions and

services based, in part, on risk will be less optimal for some groups compared to others. Although interrater reliability may increase, this may, in turn, reduce the effectiveness of services where juveniles from some populations are inappropriately, and systematically, matched to ill-fitting sanctions and services.

In this article, we describe the effects of race/ethnicity and gender on the NCAR's predictive validity. Three research questions are addressed: (a) Does the relationship between the NCAR cumulative risk score and recidivism differ by race/ethnicity and gender? (b) Does the NCAR cumulative risk score explain the association between gender and race/ethnicity with recidivism? and (c) What sources of error explain the remaining effects of race/ethnicity and gender, if any, on reoffending?

## METHOD

### PARTICIPANTS

The present study was conducted using the administrative records of all adjudicated juveniles in North Carolina from July 1, 2001, through June 30, 2003 ( $N = 14,719$ ; 48% African American, 43% White, 73% male). Administrative data were limited to risk assessment scores, ratings of the seriousness of the intake offense, dates of disposition, disposition levels, and demographic information.

The study sample consisted of African American and White juveniles who were adjudicated for delinquent offenses and who were exposed to the possibility of recidivism. These selection criteria resulted in the exclusion of 5,185 juveniles from the study. Of these, 394 juveniles were excluded because they were sent to youth development centers and were therefore not given the opportunity to reoffend. In addition, 3,095 were excluded because they were adjudicated for nondelinquent offenses, such as traffic violations, truancy, and running away. Finally, 1,205 youths classified by the juvenile court as Latina/Latino, Native American, Asian American, and biracial were also excluded, because samples sizes were too small to permit analysis.<sup>1</sup> Therefore, analyses were based on 9,534 delinquent offenders (54% African American, 23% female; mean age = 13.7 years,  $SD = 1.4$ ) who received community-based dispositions and thus had the opportunity to reoffend.

## MEASURES

Court counselors assessed risk of recidivism using the NCAR. The North Carolina Department of Juvenile Justice and Delinquency Prevention designed the NCAR in a collaborative process that included consultation with the National Council on Crime and Delinquency (NCCD) and researchers from the School of Social Work at the University of North Carolina at Chapel Hill. The NCAR follows closely the structure and format recommended by the U.S. Office of Juvenile Justice and Delinquency Prevention and is similar to many instruments developed for states and local jurisdictions by the NCCD (Howell, 1995). Court counselors completed the NCAR following intake interviews with the juvenile, the juvenile's parents, law enforcement officers, school officials, and other community providers.

The NCAR is a nine-item index of risk factors. Each risk factor includes between two and five response options, which are summed into a cumulative risk score ranging from 0 to 30. Response options are either behaviorally anchored or matters of historical record to minimize rater discretion. Individual risk factors include (a) age when first delinquent offense was alleged in a complaint, (b) number of undisciplined or delinquent referrals to intake, (c) most serious prior adjudication, (d) number of prior assaults, (e) history of run-aways from home or placement, (f) severity of known use of alcohol or illegal drugs during the past 12 months, (g) school behavior problems during the past 12 months, (h) delinquent peer associations, and (i) parental supervision. Cronbach's alpha for the NCAR is modest ( $\alpha = .65$ ), owing to its multidimensional nature. Details about the NCAR's scoring were described in Schwalbe et al. (2004).

Recidivism was defined as a subsequent adjudication for a delinquent offense. This definition is restrictive in comparison to other studies, which have used criteria such as new arrest or referral to juvenile court and results in a conservative estimate of predictive validity.

## PLAN OF ANALYSIS

Censoring complicates analyses of recidivism data. Censoring occurs in longitudinal studies when event times (e.g., subsequent adjudications) are unobserved (Collett, 1994). In the current study, two time-related processes account for most of the censoring. First,



juveniles entered the data set at different times, resulting in different follow-up lengths. For instance, a few juveniles were adjudicated just 1 day before the end of the study, whereas several were followed for more than 700 days. Second, juveniles age-out of the juvenile justice system in North Carolina when they reach their 16th birthdays. Thus, any offenses committed after their 16th birthdays fall under the jurisdiction of the adult criminal court and are therefore not recorded in the juvenile justice record. In the present study, 2,920 juveniles aged-out of the juvenile justice system during the study window. These two censoring processes yielded a sample whose median follow-up time was 384 days (range of 1 to 728 days); 90% of the study sample had follow-up times of less than 648 days.

Because follow-up times varied, event history methods, specifically the Kaplan-Meier Product Limit estimator and Cox regression, were used to answer the study questions. The Kaplan-Meier Product Limit estimator provides descriptive statistics for the dependent variable—time to recidivism—and compares differences by race/ethnicity and gender. Cox regression models the effects of race/ethnicity, gender, and risk scores on the hazard of recidivism. The hazard of recidivism is defined as the instantaneous probability of recidivism during an interval of time, which in this study was a single day.

Data for a small number of cases ( $n = 232$ , 2.4%) were missing risk assessment scores and were therefore incomplete. Juveniles without risk assessment scores did not differ from the remaining juveniles in terms of age, class of offense, disposition level, or gender. Juveniles with missing data were more likely to be African American,  $\chi^2(1, N = 9,534) = 16.74, p < .001$ , and had higher rates of reoffending, log-rank:  $\chi^2(1, N = 9,534) = 9.71, p < .01$ . Because of the small amount of missing data, listwise deletion was used in all the analyses.

## RESULTS

Table 1 presents descriptive statistics for the sample. Of the total ( $N = 9,534$ ), 35.5% were White males, 41.5% were African American males, 10.0% were White females, and 13.0% were African American females. The majority of offenders across all groups were adjudicated

**TABLE 1: Characteristics of the Sample by Gender and Race/Ethnicity**

	Full Sample (N = 9,534)	Male		Female	
		White (n = 3,387)	Black (n = 3,955)	White (n = 955)	Black (n = 1,237)
Offense class					
Misdemeanor	61%	61%	58%	70%	62%
Felony	30%	30%	33%	25%	25%
Serious felony	9%	9%	9%	5%	13%
Any prior offenses	50%	49%	49%	39%	49%
Mean cumulative risk score (SD)	7.4 (4.52)	6.9 (4.35) <sup>b,2</sup>	7.8 (4.66) <sup>a</sup>	7.3 (4.34) <sup>b</sup>	7.5 (4.50) <sup>1</sup>
Estimated recidivism rate <sup>a</sup>	.29	.24 <sup>b,2</sup>	.35 <sup>a</sup>	.23 <sup>b</sup>	.29 <sup>b,1</sup>

Note. Statistically significant group differences ( $p < .01$ ) are denoted by superscript letters and numbers.

a. Estimated recidivism rates are the inverse of the Kaplan-Meier Product limit estimator on the 650th day.

for a misdemeanor offense (e.g., simple assault, shoplifting). The next most common offense category was less serious felony offenses (e.g., breaking and entering, arson, assault on a government officer). Finally, fewer juveniles were adjudicated for the most severe felony offenses (e.g., assault with a deadly weapon, rape, murder). The average cumulative risk score ( $M = 7.4$ ,  $SD = 4.52$ ) indicated that the average risk level was on the border between low and medium risk as classified by the NCAR. Twenty-nine percent ( $n = 2,765$ ) of the juveniles were estimated to reoffend during the follow-up period.

Table 1 shows differences across gender and race/ethnicity. Higher proportions of White females had less serious intake offenses and were more likely to be first-time offenders than juveniles from the other groups. Analysis of variance (ANOVA) of the cumulative risk score was significant,  $F(3, 9531) = 22.72$ ,  $p < .001$ , and post hoc comparisons revealed that (a) scores for African American males were significantly higher than those for White males and White females and (b) scores for African American females were significantly higher than those for White females. Log-rank statistics for rates of recidivism also indicated statistically significant differences among the groups,  $\chi^2(3, N = 9,534) = 40.10$ ,  $p < .001$ . Individual

**TABLE 2: Mean Scores and Standard Deviation on Risk Items by Gender and Race/Ethnicity**

Risk Factor	Male						Female			
	Full Sample (N = 9,534)		White (n = 3,387)		Black (n = 3,955)		White (n = 955)		Black (n = 1,237)	
	M	SD	M	SD	M	SD	M	SD	M	SD
Age at first offense	.18	.39	.18	.39 <sup>b</sup>	.23	.42 <sup>a</sup>	.07	.26 <sup>d</sup>	.13	.34 <sup>c</sup>
Number of prior referrals	.24	.32	.23	.30 <sup>b</sup>	.27	.33 <sup>a</sup>	.20	.29 <sup>b,2</sup>	.25	.29 <sup>1</sup>
Most serious prior adjudication	.14	.25	.13	.24 <sup>b,1</sup>	.17	.28 <sup>a</sup>	.09	.20 <sup>b,2</sup>	.12	.23 <sup>b,1</sup>
Prior assaults	.09	.19	.08	.17 <sup>b</sup>	.10	.20 <sup>a</sup>	.06	.15 <sup>b</sup>	.10	.20 <sup>a</sup>
Runaways	.16	.37	.12	.32 <sup>b</sup>	.13	.34 <sup>b</sup>	.29	.45 <sup>a</sup>	.27	.45 <sup>a</sup>
Alcohol or drug use	.19	.33	.23	.35 <sup>b</sup>	.15	.30 <sup>c</sup>	.30	.38 <sup>a</sup>	.13	.28 <sup>d</sup>
School behavior problems	.73	.34	.71	.35 <sup>b</sup>	.76	.33 <sup>a</sup>	.71	.35 <sup>b</sup>	.76	.33 <sup>a</sup>
Delinquent peer relationships	.30	.25	.26	.23 <sup>2</sup>	.33	.26 <sup>a</sup>	.30	.24 <sup>1</sup>	.30	.27 <sup>b</sup>
Parental supervision	.23	.33	.21	.32 <sup>b</sup>	.23	.33 <sup>a</sup>	.25	.34 <sup>a</sup>	.25	.34 <sup>a</sup>

*Note.* All risk factor scores were standardized on a common metric (0 to 1) to facilitate comparisons. ANOVAs of individual risk factors by group membership were significant ( $p > .001$ ) for all risk factors. Post hoc comparisons were conducted by Tukey's test. Within each risk factor, statistically significant ( $p < .01$ ) group differences are denoted by superscripts. For example, within number of prior referrals, superscript letters show that mean scores for Black males were higher than for White males and White females, and subscript numbers show that mean scores for Black females were higher than for White females.

comparisons showed that rates of recidivism were highest for African American males, whereas recidivism rates for African American females were higher than those for both White males and White females.

Table 2 shows the means and standard deviations for the nine risk factors measured by the NCAR. Mean scores were converted to a common metric (range of 0 to 1) to facilitate comparisons across risk factors. Seven of the nine mean scores were below 0.25, and the

median scores on these risk factors were 0, indicating no risk for most juveniles. School behavior problems was one of only two risk factors that did not follow this trend. The median score for this factor was the highest score possible, with court counselors having given "serious" ratings (e.g., more than one short-term suspension, long-term suspension, 10 unexcused absences, or expelled/dropped out) to 52% ( $n = 4,930$ ) of delinquent juveniles in this study. Similarly, the peer relationships of about 76% ( $n = 7,285$ ) of all juveniles indicated at least some risk, although only 2% ( $n = 221$ ) were rated at the highest level (gang involvement).

Table 2 also shows differences by race/ethnicity and gender. Two-way ANOVAs (gender and race/ethnicity) were significant for all risk factors. In general, patterns suggested that African American males had higher risk scores than African American females, White males, or White females. Close inspection of the table, however, shows that many of the differences were small. Risk factors showing larger group differences included age at first offense (where African American juveniles were more likely to have an early offense compared to White juveniles), running away (where females were more likely to have run away from their home or placement than males), and alcohol/drug use (where White females had higher mean scores than any other group).

The next analysis compared the bivariate hazard ratios of risk factors on the hazard of recidivism separately for each risk factor. Table 3 shows the exponentiated parameter estimates for individual Cox regression models for each factor in the full sample as well as in subsamples defined by gender and race/ethnicity. The analysis showed that eight of the nine risk factors were significant predictors of the hazard of recidivism within the total sample as well as for both White and African American males. Age at first offense was not a significant predictor. In contrast, five of nine risk factors significantly predicted hazard of recidivism for African American females, whereas only one predicted hazard of recidivism for White females. Among African American females, only one of the offense history variables predicted recidivism, whereas four of the five nonoffense history factors (alcohol/drug use, school behavior problems, delinquent peers, and parental supervision) predicted hazard of recidivism. Among White females, only prior assaults achieved statistical

**TABLE 3: Hazard Ratios of Individual Risk Factors on Recidivism by Gender and Race/Ethnicity**

Risk Factor	Full Sample (N = 9,534)	Male		Female	
		White (n = 3,387)	Black (n = 3,955)	White (n = 955)	Black (n = 1,237)
Age at first offense	1.021	1.082	0.955	0.991	0.969
Number of prior referrals	1.375***	1.377***	1.390***	1.073	1.340***
Most serious prior adjudication	1.225***	1.270***	1.229***	0.971	1.035
Prior assaults	1.198***	1.230***	1.165***	1.247*	1.146
Runaways	1.241***	1.331***	1.277***	1.237	1.170
Alcohol or drug use	1.165***	1.202***	1.238***	1.053	1.199*
School behavior problems	1.277***	1.267***	1.265***	1.184	1.356**
Delinquent peer relationships	1.180***	1.190***	1.163***	1.060	1.207***
Parental supervision	1.213***	1.284***	1.185***	1.063	1.258**
NCAR total risk score	1.081***	1.097***	1.076***	1.039	1.073***

Note. NCAR = North Carolina Assessment of Risk.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

significance. One of the nonsignificant factors (runaways) approached statistical significance ( $p = .06$ ) and had a hazard ratio of similar magnitude to the total sample (1.237 vs. 1.241). Nevertheless, more than 75% of the risk factors included in the NCAR did not differentiate White female repeat offenders from nonrepeat offenders during the time period covered by this study.

Table 3 also includes an analysis of the relationship between cumulative risk score and the hazard of recidivism for the total sample, as well as separately by race/ethnicity and gender. In the total sample, the hazard ratio increases 8% for every point increase in the total risk score ( $r = .11$ ). A 5-point increase in the risk score yielded a 48% increase in the hazard ratio.<sup>2</sup> The magnitude of the effects were similar for White males and both African American males and

females, where the hazard ratio indicated a 7% to 10% increase in the hazard of recidivism for every point increase in the total risk score ( $r_{\text{Whitemale}} = .13$ ;  $r_{\text{Blackmale}} = .12$ ;  $r_{\text{Blackfemale}} = .11$ ). A 5-point increase in risk score yielded an increase in the hazard ratio from 42% to 59% for these groups.

The effect for White females was not significant ( $r_{\text{Whitefemale}} = .03$ ,  $p = .10$ ). Examination of the parameter estimates and their associated standard errors reveals not only that the effect size was smaller for White females than for the other groups but also that the standard error of the parameter estimate was higher. The coefficient of variation (CV) is a ratio of the standard error of an estimate divided by the estimate (Peck, Olsen, & Devore, 2001). It is a measure of the variability of an estimate, where smaller ratios indicate smaller degrees of variability compared to larger ratios. CVs in the present case ( $CV_{\text{Whitemale}} = .12$ ;  $CV_{\text{Blackmale}} = .11$ ;  $CV_{\text{Blackfemale}} = .24$ ;  $CV_{\text{Whitefemale}} = .60$ ) showed that the variability around the parameter estimate for White females was 3 to 5 times less precise than for the other groups. This lack of precision, accompanied by the small effect size, resulted in nonsignificant findings and suggests that the relationship between measured risk and recidivism was different for White females compared to other groups.

Tables 4 and 5 show the test for the mediation of race/ethnicity and gender effects by the cumulative risk score. Model 1 (shown in Table 4) is the final model after testing for all possible interactions and quadratic terms. It shows that parameter estimates for race/ethnicity, gender, the cumulative risk score, and a quadratic term for the cumulative risk score all achieved statistical significance. Seriousness of the instant offense, included in early model development, was not significant and was therefore dropped from the analysis. Estimate signs were positive for race/ethnicity, gender, and cumulative risk score, indicating that African American juveniles, males, and juveniles with higher NCAR scores had a higher hazard for recidivism compared to White juveniles, females, and juveniles with lower NCAR scores. The negative sign for the quadratic term indicates a nonlinear relationship between cumulative risk and the hazard for recidivism, such that increases in the hazard rate were smaller at higher levels of risk than they were at lower levels.

**TABLE 4: Cox Regression of NCAR Total Risk Score, Race/Ethnicity, and Gender on Hazard of Recidivism**

	<i>Model 1</i>		<i>Model 2</i>	
	<i>B</i>	<i>SE</i>	<i>B</i>	<i>SE</i>
African American	.26394	.05849***	.32217	.05830***
Male	.20588	.07152**	.19702	.05830***
NCAR risk score	.17787	.02267***		
NCAR risk score square	-.00516	.00111***		

*Note.* NCAR = North Carolina Assessment of Risk.

\*\* $p < .01$ . \*\*\* $p < .001$ .

**TABLE 5: Direct and Indirect Effects of Race/Ethnicity on Hazard of Recidivism**

<i>Type of Effect</i>	<i>Magnitude of Effect</i>	<i>Percentage of Effect</i>
Direct effect	.264***	82
Indirect effect (through NCAR total risk score)	.058***	18
Total effects	.322***	100

*Note.* Significance of the indirect effects calculated using procedures recommended by Clogg, Petkova, and Haritou (1995). NCAR = North Carolina Assessment of Risk. \*\*\* $p < .001$ .

Mediation is present when parameter estimates for mediated variables (race/ethnicity and gender in the present study) are substantially reduced when controls for mediating variables are introduced into the model. In this case, Model 2 (shown in Table 4) shows an initial model that included only race/ethnicity and gender, whereas Model 1 shows the full model that included the potential mediator (NCAR total risk score). Model 1 suggests a possible mediation effect, in that the parameter estimate for race/ethnicity decreased when NCAR total risk score was included in the model.

Table 5 shows the decomposition of the mediation of race/ethnicity effects. The total effect of a mediated variable is the magnitude of its parameter estimate when the mediator is omitted. The direct effect of a mediated variable is its parameter estimate when the potential mediator is included in the model. The indirect effect is the difference between these two estimates. It represents the degree to which the

mediating variable explains the relationship between an independent and dependent variable. In the present analysis, the NCAR total risk score mediated 18% of the effects of race/ethnicity on the hazard for recidivism. The indirect effect was statistically significant.

## DISCUSSION

The results of this study indicate that the predictive validity of the NCAR differed by both gender and race/ethnicity. Specifically, the NCAR risk score predicted recidivism for all groups except White females. Moreover, the NCAR risk score partially mediated the relationship between race/ethnicity and recidivism, explaining about 18% of the variation. Thus, the interpretation of NCAR scores is to some extent population dependent, but risk assessment may have the potential to reduce race/ethnicity disparities.

With respect to race/ethnicity effects, the NCAR underpredicted recidivism for African American youths, suggesting an omitted variable bias. Ideally, one might expect the NCAR risk score would explain the higher rate of recidivism for African American juveniles compared to White juveniles. That eight of nine risk factors predicted recidivism for both groups of males supports this expectation. Notwithstanding, a large proportion of the variance in recidivism attributed to race/ethnicity was unexplained. In effect, race/ethnicity remained a strong predictor of recidivism even when controlling for the effects of cumulative risk. The NCAR's brevity, which was thought to be an administrative advantage, may have unduly constrained the instrument's validity because so few predictors were used.

One strategy is to expand the scope of measured risk. Measuring risk factors that are disproportionately distributed across racial/ethnic groups may increase the sensitivity of risk assessment for diverse groups. Examples include such individual factors as conduct problems and oppositional defiant disorder (Cottle et al., 2001; McDermott & Spencer, 1997; Teplin, Abram, McClelland, Dulcan, & Mericle, 2002), socioeconomic factors such as neighborhood disadvantage (Hawkins, Laub, Lauritsen, & Cothorn, 2000; Peoples & Loeber, 1994), and sociostructural factors such as law enforcement surveillance (Brownfield et al., 2001). Directly integrating some of these



factors into risk assessment instruments and including others in multivariate models (as control variables) may help to further specify the complex relationship between risk, race/ethnicity, and recidivism.

Unlike the race/ethnicity findings for males, in which risk factors had a similar relationship to recidivism for both groups, the same did not hold true for females. This suggests problems other than omitted variable bias. According to our earlier discussion, two possibilities remain, namely, the dimensional identity problem and sampling bias. The dimensional identity problem would merit further analysis if gender differences were noted across both racial/ethnic groups. However, bivariate analyses of risk factors and recidivism (Table 3) showed that many of the risk factors predicting recidivism for males also predicted recidivism for African American females. That is, it was White females whose risk profiles differed from the others, meaning gender effects on the NCAR's predictive validity depend on race/ethnicity.

Sampling bias may account for this finding. Prior studies have identified different pathways into the juvenile justice system for female and male delinquents. Historically, gatekeepers for the juvenile justice system have been slower to refer White females to juvenile courts compared to other groups (Chesney-Lind & Shelden, 1998; Leiber & Mack, 2003; Leiber & Stairs, 1999; MacDonald & Chesney-Lind, 2001). This could affect the NCAR's predictive validity for White females in one of two ways. First, because their court referrals are timed differently, adjudicated White females may differ from other groups in meaningful ways, such as degree of family discord and history of sexual abuse. Brief risk assessment instruments such as the NCAR may not be sensitive to these gender differences. As a consequence, adjudicated White females may represent a distinct population with unique population parameters compared to the other groups. Second, selection effects could pose a methodological challenge for studies that use officially observed recidivism to estimate predictive validity. As an outcome variable, officially observed recidivism measures decision-maker tendencies along with juvenile behaviors. If White females are less likely to be referred to juvenile courts, or are referred to the courts for different reasons, then their recidivism rates will be suppressed and our

ability to estimate predictive validity will be hampered. Consequently, systematic differences in the tendency to refer female offenders will influence predictive validity studies that use this measure. Either of these hypotheses could account for the group differences observed in the present study.

These findings concerning gender suggest two recommendations for research and policy. On one hand, more sophisticated research methods that use multiple measures of recidivism are recommended. This would enable future studies to avoid problems associated with the juvenile justice selection biases suspected in the present study. On the other hand, juvenile justice researchers and practitioners committed to brief risk assessment instruments may need to evaluate the potential of separate, gender-specific instruments to promote accurate and equitable juvenile justice decision making for these groups. Although this may seem a daunting task, the present study suggests that current risk assessment practices may fail to provide valid information for some groups of female offenders.

Caution is warranted when generalizing these findings to other juvenile justice risk assessment instruments. To a certain extent, geography binds the findings of this study to the North Carolina juvenile justice context. However, the structure and content of the NCAR are similar to other brief instruments currently used across the country. Moreover, the distribution of risk factors across demographic groups was similar to other study samples in the field, which show that African American juveniles are at higher risk than White juveniles (Peeples & Loeber, 1994), that females run away with greater frequency (Funk, 1999), and that White youths have higher rates of alcohol and drug use than African American youths (Blum et al., 2000; Wagner, Lloyd, & Gil, 2002). Finally, our findings concur with emerging trends, which indicate higher levels of brief risk assessment predictive validity for White juveniles and males compared to African American juveniles and females (e.g., Schwalbe et al., 2004; Sharkey et al., 2003). Thus, this study, in concert with others, supports a broader survey of comparative predictive validity across diverse populations to determine the limits of this emergent trend and to identify conditions under which predictive validity is more similar than different.

## CONCLUSION

Ultimately, the goal of this study was to strengthen risk assessment to more fully support juvenile justice decision making. For differences across race/ethnicity, the findings suggest expanding the scope of risk factors to include factors that may predict offending for all juveniles but may be distributed disproportionately across populations. This might increase the capacity for risk assessment to mediate differences in rates of recidivism between African American and White juveniles. For differences across gender, the data suggest that innovations, such as gender-specific instruments or scoring protocols, warrant exploration in an effort to increase the validity of risk assessment classifications for females. Solutions such as these would make risk assessment classifications more precise and more helpful for juvenile court deliberation and decision making.

## NOTES

1. Subsamples (cell sizes) do not sum to 5,185 because of overlapping criteria.
2. Hazard ratio =  $\exp(z\beta)$ ,  $z$  = number of point increase in total risk score,  $\beta$  = parameter estimate for total risk score.

## REFERENCES

- Baird, C., Wagner, D., Healy, T., & Johnson, K. (1999). Risk assessment in child protective services: Consensus and actuarial model reliability. *Child Welfare, 78*, 723-748.
- Blum, R. W., Beuhring, T., Shew, M. L., Bearinger, L. H., Sieving, R. E., & Resnick, M. D. (2000). The effects of race/ethnicity, income, and family structure on adolescent risk behaviors. *American Journal of Public Health, 90*, 1879-1884.
- Brownfield, D., Sorenson, A. M., & Thompson, K. M. (2001). Gang membership, race, and social class: A test of the group hazard and master status hypotheses. *Deviant Behavior, 22*, 73-89.
- Chesney-Lind, M., & Shelden, R. G. (1998). *Girls, delinquency, and juvenile justice* (2nd ed.). Belmont, CA: West/Wadsworth.
- Clogg, C. C., Petkova, E., & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology, 100*, 1261-1293.
- Collett, D. (1994). *Modeling survival data in medical research*. New York: Chapman & Hall.
- Cottle, C. C., Lee, R. L., & Heilbrun, K. (2001). The prediction of criminal recidivism in juveniles: A meta-analysis. *Criminal Justice and Behavior, 28*, 367-394.
- Funk, S. J. (1999). Risk assessment for juveniles on probation: A focus on gender. *Criminal Justice and Behavior, 26*, 44-68.

- Hawkins, D. F., Laub, J. H., Lauritsen, J. L., & Cothorn, L. (2000). *Race, ethnicity, and serious and violent juvenile offending*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.
- Hoge, R. D. (2002). Standardized instruments for assessing risk and need in youthful offenders. *Criminal Justice and Behavior, 29*, 380-396.
- Howell, J. C. (1995). *Guide for implementing the comprehensive strategy for serious, violent, and chronic juvenile offenders*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.
- Howell, J. C. (2003). *Preventing and reducing juvenile delinquency: A comprehensive framework*. Thousand Oaks, CA: Sage.
- Ilaqu, G. E., Coulson, G. E., Lombardo, D., & Nutbrown, V. (1999). Predictive validity of the Young Offender Level of Service Inventory for criminal recidivism of male and female young offenders. *Psychological Reports, 84*, 1214-1218.
- Jung, S., & Rawana, E. P. (1999). Risk and need assessment of juvenile offenders. *Criminal Justice and Behavior, 26*, 69-89.
- Lahey, B. B., Goodman, S. H., Waldman, I. D., Bird, H., Canino, G., Jensen, P., et al. (1999). Relationship of age of onset to the type and severity of child and adolescent conduct problems. *Journal of Abnormal Child Psychology, 27*, 247-260.
- Leiber, M. J., & Mack, K. Y. (2003). The individual and joint effects of race, gender, and family status on juvenile justice decision-making. *Journal of Research in Crime and Delinquency, 40*, 34-70.
- Leiber, M. J., & Stairs, J. M. (1999). Race, contexts, and the use of intake diversion. *Journal of Research in Crime and Delinquency, 36*, 56-86.
- Lipsey, M. W., & Derzon, J. H. (1998). Predictors of violent or serious delinquency in adolescence and early adulthood. In R. Loeber & D. P. Farrington (Eds.), *Serious and violent juvenile offenders* (pp. 86-105). Thousand Oaks, CA: Sage.
- Loeber, R., & Farrington, D. P. (Eds.). (1999). *Serious and violent juvenile offenders*. London: Sage.
- MacDonald, J. M., & Chesney-Lind, M. (2001). Gender bias and juvenile justice revisited: A multiyear analysis. *Crime & Delinquency, 47*, 173-195.
- McDermott, P. A., & Spencer, M. B. (1997). Racial and social class prevalence of psychopathology among school-age youth in the United States. *Youth & Society, 28*, 387-414.
- Peck, R., Olsen, C., & Devore, J. (2001). *Introduction to statistics and data analysis*. Pacific Grove, CA: Duxbury.
- Peebles, F., & Loeber, R. (1994). Do individual factors and neighborhood context explain ethnic differences in juvenile delinquency? *Journal of Quantitative Criminology, 10*, 141-157.
- Pope, C. E., & Feyerherm, W. H. (1995). *Minorities and the juvenile justice system*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.
- Schwalbe, C. S., Fraser, M. W., Day, S. H., & Arnold, E. M. (2004). North Carolina Assessment of Risk (NCAR): Reliability and predictive validity with juvenile offenders. *Journal of Offender Rehabilitation, 40*, 1-22.
- Sharkey, J. D., Furlong, M. J., Jimerson, S. R., & O'Brien, K. M. (2003). Evaluating the utility of a risk assessment to predict recidivism among male and female adolescents. *Education and Treatment of Children, 26*, 467-494.
- Teplin, L. A., Abram, K. M., McClelland, G. M., Dulcan, M. K., & Mericle, A. A. (2002). Psychiatric disorders in youth in juvenile detention. *Archives of General Psychiatry, 59*, 1133-1143.

- Von Eye, A., & Bergman, L. R. (2003). Research strategies in developmental psychopathology: Dimensional identity and the person-oriented approach. *Development and Psychopathology, 15*, 553-580.
- Wagner, E. F., Lloyd, D. A., & Gil, A. G. (2002). Racial/ethnic and gender differences in the incidence and onset age of *DSM-IV* alcohol use disorder symptoms among adolescents. *Journal of Studies on Alcohol, 63*, 609-619.