Article

Crime & Delinquency
2014, Vol. 60(2) 167–192

Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0011128711426544
cad.sagepub.com
SAGE

An Analysis of Prisoner Reentry and Parole Risk Using COMPAS and Traditional Criminal History Measures

Sheldon X. Zhang[1], Robert E. L. Roberts[2], and David Farabee[3]

Abstract

The California Department of Corrections and Rehabilitation has adopted Correctional Offender Management and Profiling Alternative Sanctions (COMPAS), an actuarial risk- and needs-assessment instrument, as part of its reentry supervision and parole planning procedure. A large-scale 3-year prospective study was conducted to assess the instrument with regard to how well it predicted whether a parolee would be rearrested for (a) any crime and (b) a violent offense. This study followed, for up to 2 years, a total of 91,334 parolees who had been assessed with COMPAS prior to reentry into the community. The instrument achieved an acceptable level of predictive validity in general rearrests with an area under the curve value of 0.70, but its predictive power for subsequent violent offenses fell short of this conventional threshold. Moreover, a parsimonious model using four known risk factors from existing official records (i.e., gender, age, age of first arrest, and the number of prior arrests) performed just as well in predicting subsequent

[1]San Diego State University, CA, USA
[2]California State University San Marcos, USA
[3]University of California, Los Angeles, USA

Corresponding Author:
Sheldon X. Zhang, Department of Sociology, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-4423, USA.
Email: szhang@mail.sdsu.edu

arrests. Findings from this study illustrate the challenges in applying group-based attributes to predict individual criminal behavior and suggest that, although COMPAS has other attractive features such as case management capability, existing official records may offer a lower cost alternative for assessing the risk of reoffending for community reentry purposes.

Keywords
prisoner reentry, risk assessment, COMPAS, predictive validity

Background

A continual challenge in prisoner reentry is to assess inmate's risk of reoffending and their need for correctional services. As of 2009, more than 1.5 million people were incarcerated in state and federal prisons in the United States, more than 760,000 were incarcerated at the local level, and more than 7.2 million people were under some form of correctional authority supervision (Glaze, 2010). High rates of incarceration mean high volumes of prisoner reentry and, hence, competition for the allocation of scarce correctional resources. According to the Pew Center on the States (2009), as of 2007, 1 in every 31 adults in the United States (3.2% of the adult population) was under some form of criminal justice supervision, with the size of this population tripling over the course of 25 years, between 1982 and 2007 (Pew Center on the States, 2009). Risk and needs assessments therefore become an important aspect of reentry planning for correctional agencies to ensure appropriate supervision and programming in the community, such as parole supervision intensity and treatment plans.

High rates of reoffending among parolees have remained a challenge for correctional agencies at both state and federal levels. There have been two widely cited national recidivism studies conducted by the Bureau of Justice Statistics: The first was based on a 3-year follow-up of state and federal prison inmates released in 1983 (Beck & Shipley, 1989) and the second was based on a 3-year follow-up of inmates released in 1994 (Langan & Levin, 2002). More than two thirds of those released from prison in 1994 were rearrested within 3 years, representing an 8% increase over the cohort released in 1983. Recidivism rates in California are on par with national figures. More than 40% of California's offenders return to prison within 1 year, more than 50% within 2 years, and two thirds in 3 years (California Department of Corrections and Rehabilitation [CDCR], 2009). Considering the fact that California operates one of the largest prison systems in the country (a close second only to Texas) and makes about 120,000 releases to parole each year, the persistently high return-to-custody rates present serious challenges for policy makers and agency administrators alike on how best to allocate correctional resources and reduce recidivism.

One of the main efforts that the CDCR made in recent years is to incorporate actuarial risk and needs assessment to guide its supervision and community-based reentry programs. The agency began using the Correctional Offender Management and Profiling Alternative Sanctions (COMPAS) automated risk- and needs-assessment system in its prison in 2007 (CDCR, 2009). By 2008, the COMPAS instrument was adopted by all institutions as part of its prerelease planning for field supervision and referrals to correctional treatment. The use of COMPAS was later expanded to prison reception centers where incoming prisoners were classified and assigned to different institutions according to the identified risk levels and needs.[1]

The Rise of Actuarial Risk Assessment for Correctional Populations

The basic feature of actuarial risk assessment, as opposed to subjective evaluation, is the use of statistical algorithms to establish risk profiles associated with various groups of individuals that share certain characteristics. Analogous to the underwriting process of a life insurance policy, statistical methods are devised by researchers to weigh varied configurations of factors that are theoretically and empirically associated with reoffending, and then individual offenders are classified into different groups based on their shared similarities in terms of likelihood to reoffend (Rice & Harris, 2005; Silver, Smith, & Banks, 2000).

The use of actuarial risk and needs assessment has been around for a long time. According to Grove and Meehl (1996), sociologist Ernest Burgess combined 21 objective factors in a crude form of actuarial algorithm in 1928 to assess the likelihood of parole failure among 3,000 inmates released from three Illinois prisons, while three prison psychiatrists were asked to do the same with the same available information, but relying on their own clinical judgment. Burgess' actuarial prediction was found to be superior to that of the psychiatrists and was considered the beginning of a direct challenge by actuarial assessment against subjective judgment by correctional officials and clinicians (Grove & Meehl, 1996).

Recent studies and literature reviews continue to support this finding. For instance, Grove, Zald, Lebow, and Snitz (2000) compared the accuracy of clinical and statistical assessment in a meta-analysis on 136 studies of human health and behavior, and found that on average, statistical methods were about 10% more accurate than clinical predictions. In various other meta-analyses and literature reviews, the superiority for actuarial assessment is consistent in a wide range of circumstances (Andrews, Bonta, & Wormith, 2006; Mossman, 1994). These meta-analyses found that the actuarial methods outperformed clinical (or professional) judgment by rather wide margins in terms of their predictive accuracy (Andrews & Bonta, 2006).

Actuarial risk assessment has gained much notice and acceptance in the criminal justice system as a tool to facilitate decision making, despite concerns over applying group-based attributes to predict individual behavior and risk of continued marginalization of sizable groups of populations already at the fringes of the society (Silver & Miller, 2002). Although subjective judgment can still predict future offending in certain contexts (Grove et al., 2000), the use of actuarial risk-assessment tools has now become an accepted standard of forensic risk-assessment practice (Monahan et al., 2001). In most cases, actuarial tools are designed by combining empirically or theoretically derived constructs that are predictive of violence or antisocial activities to guide the forecasting of future antisocial or violent acts.

Actuarial instruments have two distinct features that appeal to correctional agencies—transparency and efficiency. Large volumes of information, from static (i.e., preexisting or constant) variables such as prior criminal history, drug abuse, education, and employment status to dynamic (i.e., contemporaneous or fluctuating) measures including attitudes and self-perceptions are quickly processed. For political and pragmatic reasons, the criminal justice system must consider the likelihood of further violent and nonviolent behavior among those brought to its attention (Andrews & Bonta, 2006). Practitioners in correctional agencies must also decide on the allocation of supervision resources, which is mostly determined by the perceived risk of reoffending and the needs of those on parole or probation.

The science of predicting criminal behavior and assessing offender correctional needs has also gone through major changes in the past 30 years, with an increasing number of justice agencies either developing in-house tools or adopting commercial instruments (Andrews & Bonta, 2006). According to Bonta (1996), risk assessment went through three major changes: (a) from subjective evaluation rendered by clinicians or criminal justice officials who relied on personal knowledge and sometimes intuition, to (b) statistical analysis of static information in offenders' case files, to (c) an emphasis on theoretically as well as empirically based measures. Recently, Andrews et al. (2006) proclaimed that a fourth generation of risk assessment had emerged, featuring not only theoretically relevant and evidence-based risk assessment but also case management that can guide correctional officials in planning and executing supervision and intervention activities.

One method of assessing the predictive accuracy of an actuarial instrument is correlation analysis, using such common procedures as Pearson product–moment correlation (Rice & Harris, 2005). In forensic science, however, the point-biserial (mathematically equivalent to the Pearson product–moment) correlation ($r_{pb,}$) is often used, because the outcome variables are often dichotomous (e.g., recidivism). The more common measure of determining the predictive accuracy of a risk-assessment instrument these days is the area under the receiver operating characteristic (ROC) curve (or area under the curve [AUC]), which plots the ratio of true positives and false positives.

The statistical analysis of AUC came from signal detection theory, first developed by researchers to quantify the ability of radar operators in World War II to interpret signals against noise (Marcum, 1947). The basic logic is to compare the rate of predictions of event *i* that prove true (true positives) to the rate of predictions of event *i* that prove false (false positive) across criterion levels of a predictive instrument. This statistical method saw an early adoption by the medical profession as a tool to screen optimal diagnostic tests (Hanley & McNeil, 1982; Lusted, 1971). More than a decade ago, researchers in forensic science began using this tool in earnest to improve the predictive accuracy of actuarial risk-assessment instruments, because of its easy interpretation and robust statistical property (Mossman, 1994; Rice & Harris, 2005). Nearly all current studies of actuarial risk-assessment instruments calculate and report AUC values, either through retrospective or prospective data collection.

AUC values range from 0.50 (i.e., entirely by chance or no predictive value) to 1.00 (i.e., 100% accurate). The greater the AUC value the more precise the risk-assessment instrument. Generally, instruments that can produce AUC values of 0.70 or above are considered acceptable for clinical application purposes, values 0.60 and 0.70 are considered low to moderate, and values above 0.80 are considered strong (Aos & Barnoski, 2003; Brennan, Dieterich, & Ehret, 2009; Quinsey, Harris, Rice, & Cormier, 1998). However, Rice and Harris (2005) set much lower thresholds for the minimum values to be considered for the different sizes of measurement effects—small at AUC = 0.556 or r_{pb} = .100, medium at AUC = 0.639 or r_{pb} = .243, and large at AUC = 0.714 or r_{pb} = .371.

In practice, however, there are no set rules regarding what specific AUC value should be accepted as the threshold for considering an instrument worthy of adoption. Even so-called conventions seem to vary from study to study, as different authors tend to attach different semantics to AUC values implying somewhat different thresholds for acceptance (Brennan et al., 2009). For example, Flores et al. (2006) considered AUC values of 0.689 in a test of the Level of Services Inventory–Revised (LSI-R) as valid and robust. Kroner and colleagues (2007) considered an AUC of 0.703 as highly predictive accuracy in a study of the Violence Risk Appraisal Guide (VRAG). In another test of a modified VRAG instrument, Harris, Rice, and Camilleri (2004) considered the obtained AUC value of 0.72 a large effect size, declaring the instrument as robust in predicting violent behavior. For this study, we set our AUC threshold at 0.70.

COMPAS

COMPAS has several modules: risk/needs assessment, criminal justice agency decision tracking, treatment and intervention tracking, outcome monitoring, agency integrity, and programming implementation monitoring. The risk-assessment component addresses four basic dimensions: violence, recidivism, failure to appear, and community failure. Offenders are classified into three categories: high, medium, and low risk, based on cut points imposed on a 10-item scale (or deciles). The needs-assessment component of the COMPAS categorizes inmates as having low, medium, or high need for services or treatment in various areas, such as substance abuse, criminal thinking, and vocational training. On administering the assessment, a computer printout is generated for each inmate, with a specified level of risk as well as a list of services that would be appropriate to address the needs.

At the time of this evaluation, both the Institutional and Parole Division within CDCR adopted the COMPAS for assessment purposes. The COMPAS instrument was used at the prison reception centers across the state where inmates with new terms were assessed prior to their prison placement and again 6 to 8 months prior to their release as part of the case planning for parole supervision (an analysis of the COMPAS needs scales can be found in Farabee, Zhang, & Yang, 2011).

Today COMPAS remains an integral part of CDCR's reentry case planning. Parole Service Associates (PSAs), under the supervision of parole agents, carry out all case file reviews, conduct inmate interviews, and input data into the COMPAS data system. These PSAs first review inmates' files and then conduct interviews with these inmates. Depending on the extensiveness of an inmate's prior histories with the justice system, the review of an inmate's file ranged from 20 min to more than 2 hr. The interview portion of the COMPAS assessment typically lasts around 30 to 45 min. Although statistical information can be obtained through a review of an inmate's file, it is often necessary for PSAs to clarify information in the official records—especially items regarding prior community corrections or juvenile justice involvement, which are typically not captured by the adult system.

On completion of the COMPAS data entry, PSAs print out needs profiles based on the algorithms programmed in the database. These printouts are then inserted in reentry case planning packages intended for the field offices. The information provided by COMPAS is then used to assist field agents in making service referrals and decisions on levels of supervision.

Early validation studies by the developers found that the COMPAS recidivism risk model for probationers achieved satisfactory accuracies, with AUCs of 0.72 and 0.74 over a 24-month outcome period (Brennan, Dieterich, & Oliver, 2006). In a pilot study in California's parole population, the COMPAS developers also found encouraging results on the psychometric properties of the instrument (Brennan et al., 2006). Using data collected from a sample of 1,077 (male n = 786 and female n = 291) soon-to-be-released inmates in California institutions, as well as from a composite norm group of 7,381 (male n = 5,681 and female n = 1,700), COMPAS developers found satisfactory scores on measures such as internal consistency, concurrent and criterion validity, and construct validity. The distributions of the basic COMPAS scale scores in the California Parole pilot sample are very similar to the distributions in the prison/parole normative data.

In a more current validation study, with a sample of participants (N = 2,328) from a probation department in an eastern state, the developers of COMPAS found that the instrument achieved satisfactory results. The instrument was able to perform rather consistently across diverse offender subpopulations in three outcome criteria, particularly for violent offenses, in which "all nine of the cells (total sample, males or females, and three models) had AUCs between .70 and .80" (Brennan et al., 2009, p. 32). These findings led COMPAS developers to conclude that the predictive accuracies of the instrument are similar to, or slightly higher than, other major instruments in this field.

Independent evaluations of COMPAS by researchers not affiliated with the company have begun to appear recently. For instance, in a head-to-head comparison of predictive validity between LSI-R and COMPAS, Fass, Heilbrun, Dematteo, and Fretz (2008), using a retrospective design with archival, known groups of offenders released into the community from New Jersey prisons between 1999 and 2002, found that both LSI-R and COMPAS scores showed inconsistent validity when used in different ethnic/racial populations. For the LSI-R, Hispanics and Caucasians were more likely to be underclassified (predicting no arrests when the participant was actually arrested) than were African Americans. For the COMPAS, African Americans were more likely to be underclassified.

This study attempted, through a prospective design, to assess the predictive accuracy of the COMPAS risk assessment on California's general parole population. The goal of this study was to assess the degree to which COMPAS risk scores could predict future recidivism among California parolees. Two key outcome measures were examined for validating the predictive power of COMPAS—(a) a subsequent arrest for any reason following release and (b) a subsequent arrest for a violent offense (operationalized as homicide, assault, sexual assault, robbery, domestic violence, and kidnapping).

Method

The central task of this study was to assess how COMPAS predicts future recidivism among California prisoners released to parole. The analytic method involved calculating the strength of association between inmates' COMPAS risk scores and the distributions of parole violations and/or returns to prison within 2 years of parole. Two distinct analytic approaches were employed. One set of analyses focused on the relationship between COMPAS risk scores and the likelihood that a parolee would commit a violation and/or be returned to prison within a specific time period. The second examined the accuracy of prediction between COMPAS risk scores and the likelihood of being arrested for any crime and arrested for a violent crime.

Data Sources

Three primary sources of data were used for this analysis. The first was a database containing COMPAS scale decile scores collected from California inmates between February 2006 and May 2009. All selected cases had completed the COMPAS Core Full Assessment and/or Core Violence Assessment. Thus, each case had valid data to allow the assessment of the Risk of Violence and Risk of Recidivism Scales.

The second data source was a CDCR database that records the timing and location of offender movements within and between prisons and parole. The key variables selected pertained to dates of parole release, dates of prison returns, principal commitment offenses, reasons for returns, and inmate/parolee background characteristics.

The third data source consisted of calculated variables derived from arrest records maintained by the California Department of Justice. New variables were created from raw arrest data by CDCR Adult Research Branch analysts and provided to the research team under a data sharing agreement. Data collected from the two CDCR databases ran through June 30, 2009.

The Study Sample

The overall sample consisted of the 91,334 offenders who had completed the COMPAS Core Full Assessment and/or Core Violence Assessment during the study period (February 2006 to May 2009). Subsamples of 60,793 and 25,009 parolees with sufficient postrelease observation times to support recidivism analyses over 1 and 2 years, respectively, were then selected from the overall sample.

Arrest records were obtained for 91,334 parolees who had received the COMPAS assessment and had scores for the Violent Recidivism Risk Scale. All but a small fraction of those parolees also had a Recidivism Risk scale score. A subsequent arrest was operationalized as an arrest for any reason after release to parole. Arrest for a violent offense was operationalized as any postrelease arrest involving against-person offenses such as homicide, assault, sexual assault, domestic violence, robbery, and kidnapping.

Statistical Analysis

Descriptive analyses were conducted to compare characteristics of the study sample with those of the general parolee population in California during the study period. Significance tests were conducted between the study sample and the larger parolee population on any substantive differences on the demographic variables.

To assess the accuracy of COMPAS, the statistical analysis focused on the association between the Recidivism Risk Decile score (i.e., the general recidivism assessment score) and a subsequent arrest for any reason and between the Violent Recidivism Risk Decile Score (i.e., the assessment of violent offending) and a subsequent arrest for a violent offense. To gain an intuitive understanding of the association between COMPAS scales and recidivism, Spearman rank-order correlations were used to reflect the ordinal nature of the decile scores in the instrument and the dichotomous outcome variables.

Following correlation analysis, the area under the ROC curve was used to examine the predictive accuracy of the COMPAS risk scales. We calculated the rate of true positive predictions to false positive predictions across each of the 10 potential cut points provided by the decile scoring. In other words, each decile score was treated, in turn, as the threshold at which the instrument would predict a future recidivism event. The number of true positives and false positives that would result at each threshold level were calculated and then plotted against each other. The predictive accuracy of the instrument is graphically represented by the plot area under the ROC curve.

Finally, logistic regression analysis is used to compare test accuracy between COMPAS risk scales and four well-known risk factors commonly associated with recidivism (i.e., gender, age, age of first arrest, and number of prior arrests). Logistic regression is a more appropriate technique (than standard multiple linear regression) for estimating the partial and complete associations among a set of predictors and a dichotomous outcome variable. Employing a logistic transformation assures that the predicted values do not fall outside the bounds of the outcome variable (in this case “0” and “1”). Coincidentally, a measure of the predictive accuracy (the power of the model to discriminate accurately between levels of the outcome across levels of the predictors)—*c*—can be calculated for any logistic regression model. This statistic is equivalent to the area under the ROC curve statistic (ranging from 0.5—corresponding no better than random accuracy—to 1.0—corresponding to perfect prediction) and thus facilitates comparisons of predictive power across the levels of this analysis (Hosmer & Lemeshow, 1980). The goal of this analysis was to examine how COMPAS compares with a parsimonious model of empirical risk factors, because correctional agencies routinely gather and store offenders’ demographic and background information. These data can be easily downloaded and modeled statistically for risk-assessment purposes.

Findings

Sample Descriptive Statistics

The following table presents the background characteristics of the entire COMPAS sample, and subsamples with 1- and 2-year postrelease observation times. For comparison purposes, the table also includes the population of prisoners released to parole during 2008. As shown in Table 1, the full COMPAS sample does not differ markedly from the 2008 parole population. The most notable differences are the following: (a) African Americans were somewhat overrepresented and Latinos were slightly underrepresented in the COMPAS sample, (b) the participants in the COMPAS sample were slightly younger on average than those in the parole population (34.5 years vs. 37.0 years, respectively), and (c) members of the COMPAS sample were more likely to have committed a violent crime. With the exception of the age at release, these marginal differences between the COMPAS parolees and the

Table 1. Demographic Characteristics of Parolee Population, COMPAS Sample, and Subsamples With Increasing Postrelease Observation Periods.

	Parole population (2008)	COMPAS sample[a]	COMPAS sample with at least 1 year observation period	COMPAS sample with at least 2 years observation period
	%	%	%	%
Gender				
Female	10.9	10.1	11.0	11.4
Male	89.1	89.9	89.0	88.6
Race/ethnicity				
African American	24.0	27.1	26.5	26.9
Latino	41.1	37.0	37.5	37.0
White	29.8	30.9	31.1	31.5
Other	5.1	5.0	4.9	4.6
Age at release				
Median	37.0	33.7	33.4	33.2
Principal commitment offense				
Property	29.4	25.8	24.5	23.7
Persons	27.0	31.8	32.6	33.6
Drugs	29.9	29.6	30.0	30.6
Other	13.7	12.8	12.9	12.2
Year of COMPAS assessment				
2006	NA	22.6	33.8	78.1
2007	NA	40.8	58.6	21.9
2008	NA	35.0	7.5	0.0
2009	NA	1.6	0.0	0.0
Number of parolees	123,665	91,334	60,793	25,009

Note. COMPAS = correctional offender management and profiling alternative sanctions.
[a]Sample includes those parolees with arrest history data and COMPAS scores for violent recidivism risk.

2008 parole population increased slightly as observation time increased over 2 years.

Table 2 presents the recidivism patterns, measured as postrelease arrests, of the COMPAS parolees with 1- and 2-year observation time frames, respectively. About 56% of the parolees were arrested during the 1st year

Table 2. COMPAS Parolee Arrest Patterns During 1st and 2nd Years following Parole Release.

Parolee status 1 year after release	Number	Percentage of subsample	Percentage of arrests
COMPAS 1 year subsample	60,793	100.0	—
No arrest	26,555	43.7	—
Any arrest within 1 year	34,238	56.3	100.0
Nonviolent offense arrest within 1 year	26,381	43.4	77.1
Violent felony arrest within 1 year	4,573	7.5	13.4
Violent misdemeanor arrest within 1 year	3,284	5.4	9.6
Parolee status 2 years after release			
COMPAS 2-year subsample	25,009	100.0	—
No arrest	7,562	30.2	—
Any arrest within 2 years	17,447	69.8	100.0
Nonviolent offense arrest within 2 years	12,170	48.7	69.8
Violent felony arrest within 2 years	3,088	12.3	17.7
Violent misdemeanor arrest within 2 years	2,189	8.8	12.5

Note. COMPAS = correctional offender management and profiling alternative sanctions.

following parole. Most of those arrests were for nonviolent offenses. In all, 13% of the parolees were arrested for a violent crime during the 1st year. The arrest figures increased during the 2nd year. About 70% of all parolees were arrested within 2 years of their release; about 21% of the parolees were arrested for a violent offense (see Table 2).

Spearman Rank-Order Correlation Analysis

Table 3 presents a matrix of Pearson rank-order correlation coefficients reflecting levels of association among the COMPAS Risk Score Deciles and types of arrest for parolees with at least 2 years of observation time.[2] The correlation between the Recidivism Risk Score Decile and arrest for any reason was .31. Correlations between the Violent Recidivism Risk Score Decile and arrests for a violent offense were .21 for the overall sample and .16 for felony violent arrests. Separate correlation analyses were also conducted for male and

Table 3. Spearman Rank-Order Correlations Between Two COMPAS Recidivism Risk Scale Scores (Deciles) and Arrest Type Within 2 Years of Release (25,009 Parolees With 2-Year Observation Period).

	Recidivism Risk Score Decile	Violent Recidivism Risk Score	Any arrest within 2 years	Arrest for violent offense within 2 years	Arrest for violent felony within 2 years
Recidivism Risk Score Decile	1.00				
Violent Recidivism Risk Score Decile	0.68	1.00			
Any arrest in 2 years	0.31	0.29	1.00		
Violent arrest in 2 years	0.19	0.21	0.34	1.00	
Violent felony arrest in 2 years	0.13	0.16	0.25	0.72	1.00
M	6.16	6.31	0.70	0.21	0.12
SD	2.78	2.76	0.46	0.41	0.33
N	24,418	25,009	25,009	25,009	25,009

Note. COMPAS = correctional offender management and profiling alternative sanctions. All two-tailed inference tests of H0: rho = 0; $p < .001$.

female parolees. There was no gender difference in correlations between the Recidivism Risk Score Decile and arrest for any reason. For male parolees, the same patterns remain between the Violent Recidivism Risk Score Decile and violent arrests, but the correlations between the COMPAS scale and violent arrests were slightly weaker for women.

Area Under the ROC Curve

To describe the power of COMPAS risk scales to predict whether a parolee would be arrested within 2 years after release, we plotted a standard ROC curve using COMPAS decile scores as cut points. The shape of the curve is determined with reference to a vertical axis reflecting the true positive prediction rate (the proportion of all recidivists who had a particular COMPAS decile score and were rearrested within 2 years) and a horizontal

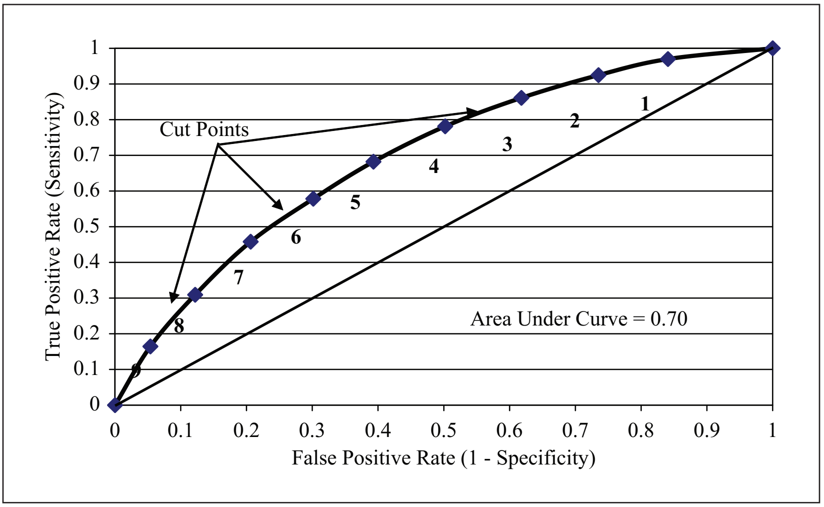

Figure 1. Receiver operating characteristic chart of Correctional Offender Management and Profiling Alternative Sanctions Recidivism Risk Scale Score Decile
Note. Outcome: Arrested within 2 years of parole release.

axis reflecting the false positive prediction rate (the proportion of all recidivists receiving a particular decile score but who did not recidivate). The AUC may vary from 0.5, a situation in which the predictive accuracy across all cut points is no better than random chance, and 1.0, a situation in which all cases are predicted accurately. A measure of the proportion of the chart AUC is provided in each figure. A diagonal line of no discrimination (AUC = 0.50) is included in each ROC figure for comparison purposes. As shown in Figures 1 and 2, both risk scales achieved levels of accuracy greater than chance, with the general recidivism scale receiving an AUC value of 0.70 and the violence scale receiving an AUC value of 0.65.

Finally, we conducted logistic regressions to compare the predictive power of COMPAS risks scales with a few well-known predictors of recidi demographic and background factors that are routinely gathered in CDCR official records, as listed in Table 1. Again, we randomly split the 2-year sample (N = 25,009) into two subsamples. One subsample was used to evaluate changes in the odds of rearrest with each increment in COMPAS Score Risk Decile. The other sample was used to estimate the predictive power achieved when relying on the demographic variables. Results of these analyses are presented in Tables 4 and 5.

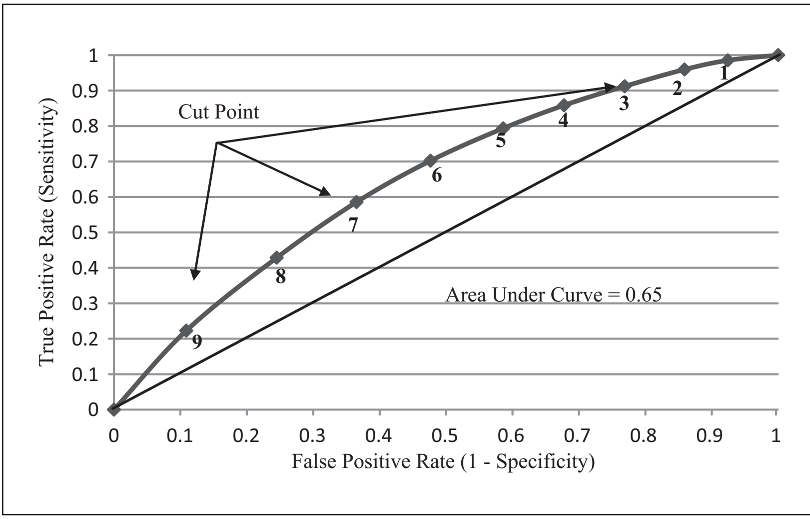

Figure 2. Receiver operating characteristic chart of Correctional Offender Management and Profiling Alternative Sanctions Violent Recidivism Risk Score Decile.
Note. Outcome: Arrested for violent offense within 2 years of parole release.

As shown under Model 1 in Table 4, the COMPAS Recidivism Risk score was a strong predictor of subsequent arrest. Each 1-unit increase in this composite score corresponded to a 30% increase in the odds of a subsequent arrest, with a test accuracy (equivalent to AUC) score of 0.70. The Model 2 includes parolee's gender, age, age of first arrest, and number of prior arrests and showed similar predictive accuracy (c = .72). The same patterns were found in the next comparison. Parallel contrasts shown in Table 6 (predicting arrests for violent offenses) again suggest that the predictive power of the well-known risk factors from CDCR's existing data yielded similar predictive accuracy (c = .67) than the COMPAS (c = .65), although neither achieved the acceptable threshold of .70.

Discussion

This study sought to assess the predictive validity of the COMPAS risk scales. Our analyses revealed that the COMPAS general recidivism and violence scales were significantly correlated with rearrests during the 24-month follow-up period, albeit moderately. The general Recidivism Risk scale was

Table 4. Parameter Estimates From Logistic Regression of Arrest Within 2 Years on COMPAS Recidivism Risk Score Decile and Parolee Characteristics.

	Model 1		Model 2		Model 3		Model 4	
	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
Intercept	2.378****	0.070	1.799****	0.074	–0.695****	0.034	0.360****	0.106
Recidivism Risk Score Decile	—	—	—	—	0.265****	0.005	0.143****	0.008
Female	–0.167****	0.022	–0.210****	0.023	—	—	–0.266****	0.024
African American	0.381****	0.039	0.342****	0.041	—	—	0.345****	0.041
Latino/a	–0.045	0.036	–0.007	0.037	—	—	–0.016	0.037
Other person of color	–0.168**	0.070	–0.056	0.073	—	—	–0.023	0.074
Age at release	–0.108****	0.002	–0.070****	–0.002	—	—	–0.050****	0.002
Age at first arrest	–0.053****	0.003	0.004	0.003	—	—	0.014****	0.003
Total prior arrests	—	—	0.086	0.002	—	—	0.061****	0.003
Test accuracy (*c*)	0.64		0.72		0.70		0.73	
Likelihood ratio χ^{2} [a]	1,425.91****		3,140.68****		2,531.16****		3,486.50****	
N	23,805		23,805		23,805		23,805	

Note. COMPAS = correctional offender management and profiling alternative sanctions.
Parolees with 24-month observation period following release.
[a]Compared with intercept-only model.
*p < .10. **p < .05. ***p < .01. ****p < .001 (two-tailed tests).

Table 5. Parameter Estimates From Logistic Regression of Violent Arrest Within 2 Years on COMPAS Violent Recidivism Risk Score Decile and Parolee Characteristics.

	Model 1		Model 2		Model 3		Model 4	
	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
Intercept	0.283***	0.086	0.071****	0.087	–2.746****	0.049	–0.955****	0.134
Violent Recidivism Risk Score Decile	—	—	—	—	0.211****	0.006	0.089****	0.008
Female	–0.277****	0.030	–0.301****	0.030	—	—	–0.337****	0.032
African American	0.267****	0.043	0.258****	0.043	—	—	0.268****	0.044
Latino/a	0.040	0.040	0.064	0.041	—	—	0.077*	0.042
Other person of color	–0.064	0.083	–0.006	0.084	—	—	0.032	0.086
Age at release	–0.033****	0.002	–0.064****	–0.003	—	—	–0.052****	0.003
Age at first arrest	–0.042****	0.004	–0.010**	0.004	—	—	0.001	0.004
Total prior arrests	—	—	0.039****	0.002	—	—	0.027****	0.002
Test accuracy (*c*)	0.64		0.67		0.65		0.67	
Likelihood ratio $\chi^{2\,a}$	1,077.03****		1,496.04****		1,201.06****		1,527.85****	
N	25,009		25,009		25,009		25,009	

Note. COMPAS = correctional offender management and profiling alternative sanctions.
Parolees with 24-month observation period following release.
aCompared with intercept-only model.
*p < .10. **p < .05. ***p < .01. ****p < .001 (two-tailed tests).

able to achieve the 0.70 AUC benchmark. However, a parsimonious model of four risk factors in existing official records produced a similar AUC value (0.72). But the predictive power of COMPAS for subsequent violent offenses fell short of the conventional threshold, yielding an AUC value of 0.65. The same model of four risk factors was again able to yield a slightly higher but still inadequate AUC value of 0.67. In separate statistical exercises, we also applied reincarceration for different reasons as outcome measures to double-check these findings for the same study sample (e.g., reincarcerated for parole violations or new convictions). The same findings remained.

Before we reflect on the study findings, we need to point out that one obvious limitation in this study is the use of official records as the outcome variables to validate the risk scales. Rearrests are more reflective of police activities than of the offender's actual criminal involvement. In other words, official records are an imprecise proxy for actual criminal activity, which COMPAS was designed to predict. Parolees are typically caught and arrested for only a small fraction of the crimes/violations they actually commit. As a result, there is a substantial amount of error in the primary outcomes used in these risk-prediction models, which likely reduces their predictive power. This is a problem that all actuarial assessment tools encounter.

Allegiance Effect and the Evaluation of Risk-Assessment Instruments

Although concordant, this study was unable to replicate the strong predictive validity of COMPAS as reported by the authors of the instrument in their own validation, which often exceeded 0.70 and sometimes 0.80 in AUC values (see Brennan et al., 2009). For instance, when used on a New York Probation sample, COMPAS was able to achieve an AUC of 0.71 in its predictive accuracy on risk of violence (Brennan, Dieterich, & Ehret, 2007). The causes of the discrepancy in effect sizes obtained by the two teams of researchers are unclear. One possible explanation could be the so-called *allegiance effect*—the fact that developers typically find greater effect sizes when validating their own instruments than those found by independent researchers (Blair, Marcus, & Boccaccini, 2008).

The allegiance effect is not unique to the development of actuarial risk-assessment tools. Similar findings have long been noted in psychotherapy literature when developers of novel intervention strategies typically found results superior to those yielded by independent studies (Luborsky et al., 1999). A recent review of the National Registry of Evidence-Based Programs and Practices, managed and funded by the Substance Abuse and Mental Health Services

Administration to assist community agencies in selecting intervention programs for their particular populations, found that much of the empirical evidence in support of the registered programs was produced and submitted by the program developers themselves (Wright, Zhang, & Farabee, 2010).

In a meta-analysis of evaluation studies on three widely known risk scales, the VRAG (Harris, Rice, & Quinsey, 1993), the Sex Offender Risk Appraisal Guide (Quinsey et al., 1998), and the Static-99 (Hanson & Thornton, 1999), Blair et al. (2008) found "compelling" evidence that an allegiance effect was key to account for the significantly larger effect sizes produced by the authors of these three instruments than those found by independent researchers. Blair et al. (2008) could find no other moderating explanations, including study design, participant characteristics, study location, and lengths of follow-up periods.

Higher measurement effects are perhaps to be expected because the developers of an assessment instrument are more adept at administering and scoring their own instrument and astute in maximizing the effect size in their statistical algorithms. Another explanation is that the instrument developers collect information specific to their computation requirement or thematic articulation, thus ensuring high levels of fidelity in the required data for analytical purposes. Furthermore, as Thase (1999) suggested in defense of the allegiance effect, the higher treatment effects are a reflection of the developer's expertise in the novel measurement or intervention approach and conversely the lower credibility and integrity of the comparator condition. In other words, when a novel instrument or therapeutic approach is applied in an environment nonspecific to, or different from, the original condition, lower effect sizes are to be expected. From an applied point of view, as Blair et al. (2008) pointed out, any expertise in achieving optimal predictive validity must be applicable across different jurisdictions and adaptable to nonspecialists who manage the day-to-day operation of the targeted populations. Otherwise, the utility of an otherwise valid instrument becomes questionable.

Challenges in Predicting Violent Behavior

Although actuarial risk assessment is superior to clinical judgment, the science of predicting actual criminality, particularly violent behavior, is rather imprecise (Dolan & Doyle, 2000). The development of a risk-assessment instrument typically uses samples in which the outcome variable is known, similar to the development of a medical diagnostic test where the disease status of a sample is known. However, a risk assessment must be applied to individuals for whom the true propensity to reoffend is unknown. Literature reviews and meta-analyses published in recent years have consistently produced mixed findings on actuarial assessment tools, which in most cases produced only moderate effect sizes. This is particularly the case in predicting violent behaviors. As shown in this study, COMPAS did not yield an adequate level of precision in predicting future violence. The studies conducted by COMPAS developers also pointed to lower AUC values in its violence scales than those found in other risk scales (i.e., general recidivism, failure to appear, or community noncompliance; Brennan et al., 2007). In another meta-analysis of nine risk-assessment tools, Min, Wong, and Coid (2010) found that all instruments were roughly the same with only moderate accuracy and essentially interchangeable in predicting future violence. They cautioned that, because these risk-assessment tools could only achieve moderate accuracy in predicting future violence, they not be used as a sole factor in justice decision making (Min et al., 2010).

There are several reasons as to why AUC values are generally lower in the prediction of violent behavior. Violence is often treated as if violent offenders all share some common traits that can be used to predict future reoffending. This may not be the case, as violent offenses may result from a wide range of factors, and each subtype seems to be precipitated by different emotive and environmental factors. For instance, resistance during robberies significantly increases the chance of violence and injuries, spousal abuse only occurs when the abuser comes in contact with his partner, and the risk of assaultive behavior increases under the influence of alcohol. Monahan and his colleagues (2001) reviewed some 134 risk factors related to violent behaviors, some of which are static while others require clinical assessment. These researchers concluded that an interactional approach is required to adequately assess the iterations of risk factors as they combine to influence the likelihood of a violent outbreak. Some variables may prove significant predictors of violence for some individuals but inconsequential to others.

Sexual violence, for instance, has been extensively studied because of the grave consequences of recidivism, and multiple instruments devised for supervision as well as treatment purposes (Barbaree, Langton, & Peacock, 2006; Bartosh, Garby, Lewis, & Gray, 2003). Although empirical studies generally support the utility of the majority of these instruments on risk assessment of sexual reoffending, one consistent finding is the marked differences in predictive accuracy among subgroups of offenders (Barbaree, Seto, Langton, & Peacock, 2001). A recent prospective study of five risk-assessment instruments commonly used for sex offenders (i.e., Static-99, Rapid Risk Assessment for Sexual Offense Recidivism, Sex Offender Risk Appraisal Guide, Sexual Violence Risk-20, and Psychopathy Checklist–Revised) yielded wide variations in their predictive accuracy (Rettenberger, Matthes, Boer, & Eher, 2010). Although the results support the utility and predictive validity of the instruments, Rettenberger and his colleagues (2010) found that none of the instruments demonstrated adequate predictive validity for sexual violent reoffending. These researchers called for particular attention to variable predictive accuracy for different sexual offender subgroups and specific offense categories.

The Need to Improve COMPAS

Predicting the risk of reoffending is an integral part of the criminal justice system, as police, judges, and correctional officials must decide what to do with offenders in sentencing, probation, or treatment activities. Correctional officers, working inside prisons or supervising field caseloads, must classify offenders into different groups based on their likelihood of reoffending. The allocation of resources and resulting supervision activities, such as surveillance, intervention, and treatment in parole or probation all require some forms of assessment of risk and needs. Such assessment bears consequences that are not trivial for the offenders as well as the larger society (Andrews & Bonta, 2006).

Although the development of actuarial measures to predict risk of reoffending has gone a long way, improving the margins of such predictions over static variables has remained a challenge (Andrews, Bonta, & Wormith, 2006). As found in this study, our parsimonious model of four common static variables was able to perform at least as well as the rather complex COMPAS scales that are considered the latest generation of actuarial risk assessment. This is not to put the COMPAS at fault but to point out the challenges in predicting individual behavior using scales derived from group attributes. The need for a standardized instrument is obvious and politically necessary for many justice agency administrators, but static variables in existing official data appear to hold their own in risk profiling.

It is also important to point out that AUC values, which are important measures of predictive accuracy in diagnostic instruments, are just one of many characteristics when evaluating a risk-assessment tool (Andrews & Bonta, 2006). There are other compelling reasons for actuarial risk-assessment instruments, aside from its predictive accuracy, such as transparency and ethical/legal considerations in criminal justice decision making (Andrews & Bonta, 2006). More importantly, the process of assessing risk through systematic and standardized measurement allows justice agencies and researchers alike to increase empirical knowledge and improve future decision making. COMPAS is more than just a risk-assessment tool. It also identifies the needs underlying the risk profiles of the assessed parolees, which are beyond the scope of this article (see Farabee et al., 2011). While the risk assessment assigns parolees into different levels of likelihood for reoffending, the assessment of needs can guide parole agents for community programming purposes. More importantly, COMPAS provides automated scoring and needs profiling whereby offender background information is summarized and graphed to facilitate field-supervision activities. The computer printout is easier to use compared with the paper and computerized records that agents must access to render their own judgments for supervision and programming decisions. However, at this point, we believe the amount of time and effort required of parole agents and their assistants to compile both static and dynamic variables for risk-assessment purposes is not cost effective. A much simpler statistical model using existing official records can accomplish the same goal with minimal staff involvement.

Authors' Note

The authors are solely responsible for the viewpoints expressed in this article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Funding for this study was provided by the California Department of Corrections and Rehabilitation.

Notes

1. It should be noted, however, that California Department of Corrections and Rehabilitation (CDCR) no longer uses the Correctional Offender Management and Profiling Alternative Sanctions to assess recidivism risk. In 2009, CDCR began using the California Static Risk Assessment, developed by Susan Turner and colleagues at the University of California, Irvine (Turner, Hess, & Jannetta, 2009).
2. A similar analysis focused on parolees with 1 year of observation time yielded similar results. Therefore, only the results for the 2-year period are presented here.

References

Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct* (4th ed.). Albany, NY: Anderson Publishing.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or needs assessment. *Crime & Delinquency, 52*, 7-27.

Aos, S., & Barnoski, R. (2003). *Washington's offender accountability act: An analysis of the Department of Correction's risk assessment*. Washington, DC: Washington Institute for Public Policy. Retrieved from http://www.wsipp.wa.gov/rptfiles/03-12-1202.pdf

Barbaree, H. E., Langton, C. M., & Peacock, E. J. (2006). Different actuarial risk measures produce different risk rankings for sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 18*, 423-440.

Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior, 28*, 490-521.

Bartosh, D. L., Garby, T., Lewis, D., & Gray, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and Comparative Criminology, 47*, 422-438.

Beck, A. J., & Shipley, B. (1989). *Recidivism of prisoners released in 1983* (Bureau of Justice Statistics Special Report NCJ 116261). Washington, DC: U.S. Department of Justice. Retrieved from http://bjs.ojp.usdoj.gov/content/pub/pdf/rpr83.pdf

Blair, P. R., Marcus, D. K., & Boccaccini, M. T. (2008). Is there an allegiance effect for assessment instruments? Actuarial risk assessment as an exemplar. *Clinical Psychology: Science and Practice, 15*, 346-360.

Bonta, J. (1996). Risk-needs assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18-32). Thousand Oaks, CA: Sage.

Brennan, T., Dieterich, W., & Ehret, B. (2007). *Research synthesis reliability and validity of COMPAS*. Traverse City, MI: Northpointe Institute for Public Management. Retrieved from http://www.northpointeinc.com/pdf/reliability_validity.pdf

Brennan, T., Dieterich, W., & Ehret, B. (2009). Evaluating the predictive validity of the COMPAS risk and needs assessment system. *Criminal Justice and Behavior, 36*, 21-40.

Brennan, T., Dieterich, W., & Oliver, W. (2006). *California Department of Corrections, Parole and Community Services Division: COMPAS Pilot Psychometric Report*. Traverse City, MI: Northpointe Institute for Public Management.

California Department of Corrections and Rehabilitation. (2009). *California Prisoners & Parolees, 2009*. Sacramento, CA: Author. Retrieved from http://www.cdcr.ca.gov/

Dolan, M., & Doyle, M. (2000). Violence risk prediction: Clinical and actuarial measures and the role of the Psychopathy checklist. *British Journal of Psychiatry, 177*, 303-311.

Farabee, D., Zhang, S. X., & Yang, J. (2011). A preliminary examination of offender needs assessment: Are all those questions really necessary? *Journal of Psychoactive Drugs, 43*, 1-7.

Fass, T. L., Heilbrun, K., Dematteo, D., & Fretz, R. (2008). The LSI-R and the COMPAS validation data on two risk-needs tools. *Criminal Justice and Behavior*. Advance online publication. doi:10.1177/0093854808320497

Flores, A. W., Lowenkamp, C. T., Smith, P., & Latessa, E. J. (2006). Validating the level of service inventory -revised on a sample of federal probationers. *Federal Probation, 70*(2), 44-48.

Glaze, L. (2010). *Correctional Populations in the United States, 2009*. Washington, DC: Bureau of Justice Statistics Bulletin, U.S. Department of Justice.

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

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19-30.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*, 29-36.

Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders* (User Report No. 1999-02). Ottawa, Ontario: Department of the Solicitor General of Canada.

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

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

Hosmer, D. W., & Lemeshow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics–Theory and Methods, 9*, 1043-1069.

Kroner, D. G., Stadtland, M., Eidt, M., & Nedopil, N. (2007). The validity of the Violence Risk Appraisal Guide (VRAG) in predicting criminal recidivism. *Criminal Behavior and Mental Health, 17*, 89-100.

Langan, P. A., & Levin, D. J. (2002). *Recidivism of prisoners released in 1994*(Bureau of Justice Statistics Special Report–NCJ 193427). Washington, DC: Bureau of Justice Statistics, U.S. Department of Justice. Retrieved from http://bjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&;iid=1134

Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., . . . Schweizer, E. (1999). The researcher's own therapy allegiances: A "wild card" in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice, 6*, 95-106.

Lusted, L. B. (1971). Signal detectability and medical decision making. *Science, 171*, 1217-1219.

Marcum, J. I. (1947). *A statistical theory of target detection by pulsed radar*. Santa Monica, CA: Rand Corporation. Retrieved from http://www.rand.org/pubs/research_memoranda/RM754.html

Min, Y., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136*, 740-767.

Monahan, J., Steadman, H. J., Silver, E., Appelbaum, P. S., Robbins, P. C., Mulvey, E. P., . . . Banks, S. (2001). *Rethinking risk assessment: The MacArthur study of mental disorder and violence*. New York, NY: Oxford University Press.

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

Pew Center on the States. (2009). *One in 31: The long reach of American corrections*. Washington, DC: Pew Charitable Trusts.

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

Rettenberger, M., Matthes, A., Boer, D. P., & Eher, R. (2010). Prospective actuarial risk assessment: A comparison of five risk assessment instruments in different sexual offender subtypes. *International Journal of Offender Therapy and Comparative Criminology, 54*, 169-186.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior, 29*, 615-620.

Silver, E., & Miller, L. L. (2002). A cautionary note on the use of actuarial risk assessment tools for social control. *Crime & Delinquency, 48*, 138-161.

Silver, E., Smith, W. R., & Banks, S. (2000). Constructing actuarial devices for predicting recidivism: A comparison of methods. *Criminal Justice and Behavior, 27,* 733-764.

Thase, M. E. (1999). What is the investigator allegiance effect and what should we do about it? *Clinical Psychology: Science and Practice, 6*, 113-115.

Turner, S., Hess, J., & Jannetta, J. (2009). *Development of the California Static Risk Assessment Instrument (CSRA)*. Irvine, CA: Center for Evidence-Based Corrections, University of California. Retrieved from http://ucicorrections.seweb.uci.edu/sites/ucicorrections.seweb.uci.edu/files/CSRA%20Working%20Paper_0.pdf

Wright, B. J., Zhang, S. X., & Farabee, D. (2010). A squandered opportunity? A review of SAMHSA's National Registry of Evidence-Based Programs and Practices for Offenders. *Crime & Delinquency*. Advance online publication. doi:10.1177/0011128710376302

Author Biographies

Sheldon X. Zhang is professor of sociology at San Diego State University. His main research interests involve community corrections, program evaluation, and transnational Chinese organized crime. He has authored three books and his articles have appeared in *Criminology, Research in Crime and Delinquency, British Journal of Criminology, Criminology* and *Public Policy, and Crime and Delinquency.*

Robert E. L. Roberts is professor of sociology at California State University, San Marcos. His main research interests are in program evaluation, recidivism risk assessment, and the social psychology of intergenerational relationships. His work has appeared in *Social Psychology Quarterly, The Journals of Gerontology, The Sociological Quarterly,* and *Journal of Marriage and the Family,* and *Crime and Delinquency.*

David Farabee is professor of psychiatry and biobehavioral sciences at the University of California, Los Angeles and a principal investigator at the Integrated Substance Abuse Programs (ISAP). He has published in the areas of substance abuse, adult and juvenile crime, HIV/AIDS, and offender treatment, was co-editor of the books *Treatment of Drug Offenders* (2002; New York: Springer) and *Treating Addicted Offenders: A Continuum of Effective Practices*, Volumes I and II (2004, 2007; New York: Civic Research Institute), and author of *Rethinking Rehabilitation: Why Can't We Reform Our Criminals?* (2005; Washington, D.C.: AEI Press).