COMPAS Validation Study: First Annual Report

December 31, 2007

Principal Investigators
David Farabee, Ph.D. (UCLA)
Sheldon Zhang, Ph.D. (SDSU)

Semel Institute for Neuroscience and Human Behavior
University of California, Los Angeles
1640 S. Sepulveda, Suite 200
Los Angeles, California 90025

California Department of Corrections and Rehabilitation
Agreement No: C06.417

UCLA
University Of California, Los Angeles

ISAP
Integrated Substance Abuse Programs

Preface

The UCLA Integrated Substance Abuse Programs (UCLA ISAP) is operating under an agreement with the California Department of Corrections and Rehabilitation (CDCR) to conduct an evaluation of the Correctional Offender Management and Profiling Alternative Sanctions (*COMPAS) assessment with regard to its ability to identify treatment needs and to predict various types of recidivism risk. Through a subcontract, UCLA ISAP is working closely with Drs. Sheldon Zhang and Bob Roberts at San Diego State University (SDSU) in this effort. The purpose of this report is to describe the results of the first year of the COMPAS evaluation.*

Table of Contents

COMPAS VALIDATION STUDY: FIRST ANNUAL REPORT

The purpose of this evaluation is to assess the implementation of COMPAS, the validity of the *needs* scales, and the predictive validity of the *risk* scales. The current report begins with a brief account of the background of the COMPAS and its use in California, followed by a discussion of implementation issues, a summary of preliminary results regarding the predictive efficacy of the risk scales, and set of updated analyses on how the COMPAS might be modified to improve its predictive accuracy among California parolees.

I. Background

COMPAS (Correctional Offender Management and Profiling Alternative Sanctions) is a computerized database and analysis system that helps criminal justice practitioners determine the placement, supervision, and case-management of offenders in community and secure settings. The data system includes several modules: risk/needs assessment, criminal justice agency decision tracking, treatment and intervention tracking, outcome monitoring, agency integrity, and programming implementation monitoring. Currently, the California Department of Corrections and Rehabilitation (CDCR) has only adopted the risk assessment component, which addresses four basic dimensions: violence, recidivism, failure to appear, and community failure. Offenders are classified into three categories of risk: high, medium, and low.

Although a recent pilot study conducted by the developers of COMPAS found encouraging results on the psychometric properties of the instrument,[1] one crucial property of the COMPAS was not adequately addressed in the COMPAS pilot study—predictive validity. Because it is expensive and time consuming to follow a sample over time to determine whether the scale can predict the likelihood of recidivism (or whether it is significantly associated with behavioral outcomes), the researchers instead performed an analysis on its concurrent validity—whether risk measures are correlated with past behaviors such as prior arrests.

The California Department of Corrections and Rehabilitation has contracted with the University of California, Los Angeles (UCLA) and San Diego State University (SDSU) to conduct an extensive evaluation of both the implementation of the COMPAS and its predictive efficacy in identifying individual risk profiles related to recidivism and treatment needs. This evaluation spans a 3-year period. This report summarizes findings from the first year of the evaluation. It should be noted that at the time the analyses were conducted for this report, only a (relatively) small number of parolees (N=515) had completed the COMPAS while incarcerated *and* had been at risk in the community for at least one year. Consequently, it would be premature to draw any conclusions regarding the validity of the instrument with California's overall parole population.

II. Implementation Issues

To gain a better understanding of COMPAS implementation issues, Drs. Farabee and Zhang conducted site visits to CIM (Chino) and R.J. Donovan (San Diego) prisons on September 27-28,

[1] Brennan, T., W. Dieterich, & W. Oliver. (2006). California Department of Corrections, Parole and Community Services Division: COMPAS Pilot Psychometric Report. Traverse City, MI: Northpointe Institute for Public Management, Inc. http://www.northpointeinc.com/.

2007. The issues raised in this section of the report are based on direct observation of the COMPAS assessment process and input from the CDCR staff responsible for carrying it out.

The COMPAS assessment process begins as early as 240 days prior to release and at minimum 120 days, at which time the C-File is requested for review and data extraction. Inmates eligible for COMPAS include those serving new terms and violators with at least six months of time in custody. CCCMS and INS holds are excluded.

Time Needed to Review C-Files

The review of the C-Files and interviewing of inmates are carried out by teams of Parole Service Associates (or PSAs) at each of the institution. The amount of time required to review C-Files varies according to the length of the inmates' criminal justice histories. Although staff reported that these reviews take an average of 45 minutes, they ranged from 20 minutes to more than two hours.

Time Needed to Interview Inmates

The interview portion of the COMPAS assessment typically lasts around 30-45 minutes. Although the first 33 questions rely mostly on information from the C-File, it is often necessary for interviewers to clarify information in the official records—especially items regarding prior community corrections or juvenile justice involvement (e.g., items 8, 10, 12, 13, 14, 17, 25, 26, 27, 28, 32, and 33).

Perceived Benefits of COMPAS

The COMPAS administrators interviewed indicated that they believed that this process significantly improves parole supervision planning. Furthermore, the PSAs indicated that because they are not viewed as correctional officersthe inmates tend to be more open in responding to their questions. As a result, PSAs contend that they are able to elicit a broad range of critical information to inform re-entry planning. Such information is vital in identifying potential problems for supervision, such as homelessness, unemployment, treatment arrangements, and transitional housing.

Another perceived benefit is the efficiency by which information is transferred from the institution to the field parole units. Because the C-File reviews and file extractions have already been conducted prior to the inmates' release, the agent of record (AOR) can quickly access the summary information in his first meeting with the parolee. Such preplanning was not possible in the previous system when the AOR had very limited time and information about a parolee he was about to supervise. The information gathered through COMPAS far exceeds that of Form CDC611 (the existing pre-release form).

COMPAS also provides case planning for inmates to learn where to get help such as driver's license, employment training, temporary living arrangement, and outpatient treatment providers. At the exit interview, PSAs provide inmates with detailed recommendations regarding services available, operating hours, living arrangements, etc.

The most important benefits of the COMPAS according to PA IIs are (1) the empirical backing of their decisions regarding supervision and referrals, and (2) the fact that—when necessary—COMPASS scores can be overridden by the parolees' AORs and/or unit supervisors. Because COMPAS does not immediately factor in changes in agency policies, unit supervisors or agents themselves must make such adjustments to actual supervision guidelines.

Perceived Problems with COMPAS

The PSAs interviewed in the course of this evaluation indicated that a number of the COMPAS interview questions are confusing to the inmates. Several of the questions are expressed as double negatives which the inmates find difficult to understand. As a result, the interviewers are forced to choose between rephrasing the questions to express their intent and strictly adhering to the interview protocol at the risk of collecting unreliable responses.

Although the service provider information in the exit interviews is perceived to be a valuable feature of the COMPAS assessment (prerelease planning), the database reflecting available services in different geographical locations must be expanded and updated on a continuous basis to remain current and relevant to parole supervision. Currently state-funded services are few and parole agents often rely on their own informal networks of services. Consequently, a more systematic and uniform database of parolee-relevant services must be developed to meet the evolving needs of the supervising agents.

The last potential problem identified during the site visits related to the uniformity in which the COMPAS assessment is administered. The interviewers vary considerably in their interviewing styles. According to the staff interviewed for this evaluation, PSAs have not received any training on interviewing techniques by CDCR, though all recognize that such training is critical to ensure greater consistency in how the data are collected.

III. Predictive Validity of Original Risk Scales

On September 4, 2007, a preliminary report on the predictive validity of the COMPAS risk scales was submitted to CDCR. The results of that report are summarized in this section. Section IV of this report builds on these findings to demonstrate how certain modifications can be made to the existing models to improve predictive validity.

It should be reiterated that the analyses described in this report are based on a preliminary sample. Table 1 (see Appendix) shows the correlations between the four major COMPAS risk scales (violence, recidivism, failure to appear, and community non-compliance), as well as their correlations with being returned to custody, having a technical parole violation, and having a non-technical parole violation. All four of the risk scales are significantly correlated with at least one of the three outcome measures. The strongest relationship was between the recidivism risk decile scores and being returned to custody within 365 days of release ($r=.27$, $p<.001$).

To assess the overall accuracy of the COMPAS risk scales, we computed Receiver Operating Characteristic (ROC) curves for the outcomes of interest. The ROC curve has become a primary measure of predictive accuracy in the past decade or so.[2] This approach involves plotting a test's rate of true positives against its rate of false positives. These curves are summarized using the area-under-the-curve (AUC) statistic, which represents the probability that if one were to randomly select a positive and negative observation (i.e., a parolee who recidivated within 12 months of release and

[2] Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, 62, 783-792. M.E. Rice & G.T. Harris. 2005. Comparing Effect Sizes in Follow-Up Studies: ROC Area, Cohen's d, and r, *Law and Human Behavior* 29(5), 615-620.

another who did not) the test would assign a higher risk score to the former than to the latter. Accordingly, an AUC of .50 would indicate that a test performs no better than chance. The greater the score deviates from .50, the better its predictive power. An AUC of 1.0 would indicate that a test perfectly classifies observations every time. In the psychological and criminological literatures, a test typically must show an AUC of .70 to be considered clinically useful.[3]

Because our analyses were hampered by the small subset of COMPAS-assessed offenders who have been at risk in the community for at least one year, we only had enough cases to roughly gauge two types of outcomes: return to prison (for any reason) and commitment of a non-technical violation of the terms of parole. As shown in Table 2, only 10% of the sample had received a technical violation. Furthermore, only 3% of the sample (n=16) had been returned for committing a violent crime (not shown).

Table 2. *Breakdown of Parolee Transitions during One-Year Observation Period*

Parolee Status One Year after Release	Number	Percent of Sample	Percent of Violation Type
Continuous Parole—No Return to Custody	261	50.7	-----
Returned to Custody	254	49.3	-----
Total Sample	515	100.0	-----
Had Technical Violation	52	10.1	100.0
Returned for Technical Violation	48	9.3	92.3
Had Non-Technical Violation	247	48.0	100.0
Returned for Non-Technical Violation	217	42.1	87.9

For each of the two outcomes described below, we present (1) a histogram showing recidivism percentages by COMPAS decile scores, and (2) a ROC curve which, as described earlier, plots a test's rate of true positives against its rate of false positives.

Return to Prison

Figure 1 shows that the COMPAS scale designed to predict overall recidivism shows a positive, linear relationship with the likelihood of being returned to prison, with only about 10% of those with a decile score of "1" being returned to custody within a year versus over 70% of those with a score decile of "10." As shown in Figure 2, this scale has an AUC value of .67.

[3] Tape, T.G. 2003. *Interpreting Diagnostic Tests*. The Area under the ROC Curve, Omaha: University of Nebraska Medical Center, see: http://gim.unmc.edu/dxtests/roc3.htm. Washington State Institute for Public Policy. 2006. Sex offender sentencing in Washington State: Predicting recidivism based on the LSI-R. http://www.wsipp.wa.gov/rptfiles/06-02-1201.pdf.

Figure 1. *Percentage of Parolees Returned to Custody within One Year by COMPAS Recidivism Risk Score Decile*

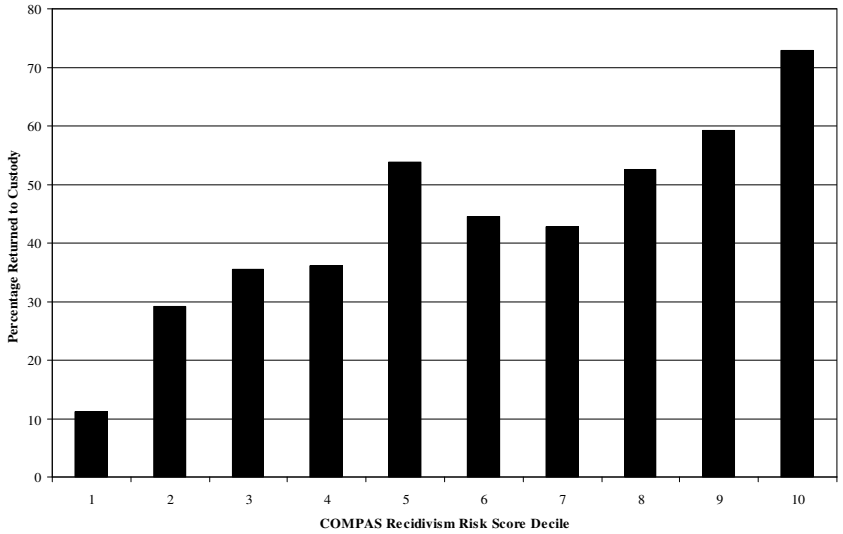
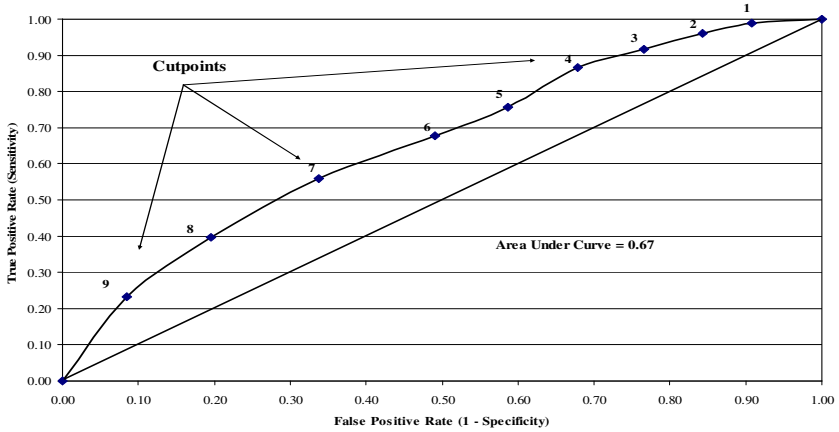


Figure 2. *ROC Chart of COMPAS Recidivism Risk Score Decile (Outcome: Returned to Custody within One Year of Parole Release)*



Non-Technical Violations

The histogram shown in Figure 3 shows a more complex relationship between the COMPAS decile scores (predicting community non-compliance) and the percentage of parolees who committed a non-technical violation.[4] As shown in Figure 4, the Community Non-Compliance scale has an AUC value of .61.

Figure 3. *Percentage of Parolees Committing Non-Technical Parole Violation within One Year by COMPAS Community Non-Compliance Risk Score Decile*

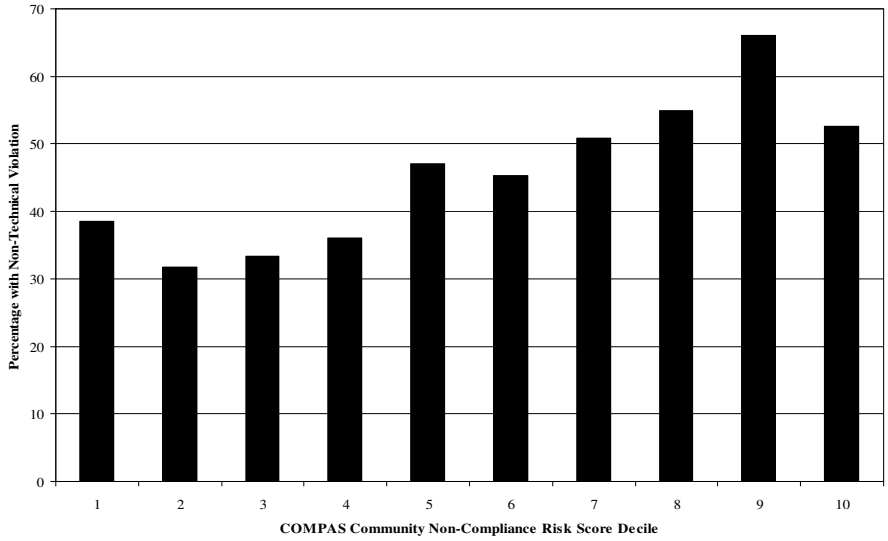
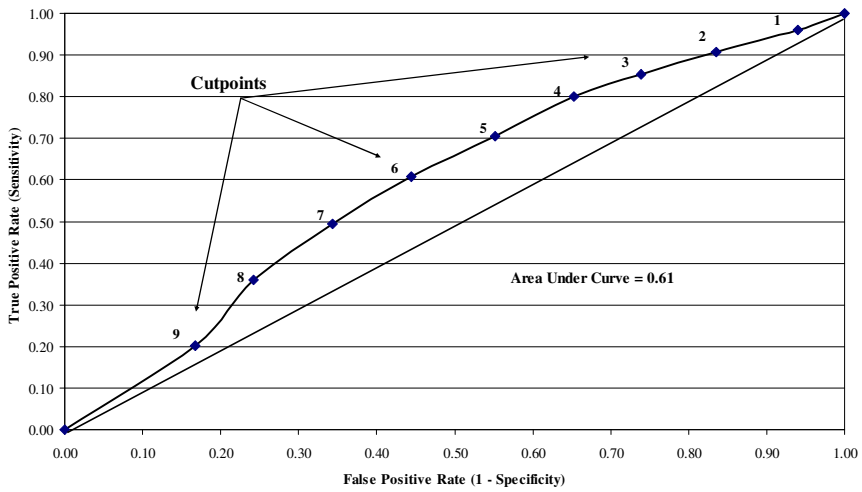


Figure 4. *ROC Chart of COMPAS Community Non-Compliance Risk Score Decile (Outcome: Non-Technical Parole Violation within One Year of Parole Release)*



[4] This outcome was not limited to parolees who were returned to custody for violating a condition of parole, though (as shown in Table 3) nearly 9 out of 10 parolees with non-technical violations were returned to prison.

IV. Predictive Validity of Modified Risk Scales

To explore possible ways by which the existing COMPAS prediction models might be improved, we first conducted a series of logistic regressions consisting of (1) standard predictors available from OBIS (Model 1), (2) all four COMPAS risk scales (Model 2), (3) the full combination of the independent variables in Models 1 and 2 (Model 3), and (4) a reduced set of these predictors based on the most parsimonious combination (Model 4). The goal of logistic regression is to accurately predict a dichotomous outcome for individual cases using the most parsimonious model. To do this, a model is created that includes all predictor variables that are useful in predicting the response variable.

The odds ratio (OR) for the independent variables gives the *relative* amount by which the odds of the outcome increase (OR > 1) or decrease (OR < 1) when the value of the independent variable is increased by 1 unit. For example, in Table 3, the variable "Number of Prior Incarcerations" has an odds ratio of 1.12, indicating that each prior incarceration term is associated with a 12% increase in the odds of being returned to custody within one year.

Table 3. Odds-Ratios from Logistic Regression of Return to Custody within One Year on COMPAS Risk Measures and Parolee Characteristics (Males only, N = 457).

Predictor	Model 1	Model 2	Model 3	Model 4
Failure-to-Appear Risk Decile	---	1.06	1.00	---
Violence Risk Decile	---	0.99	0.99	---
Community Non-Compliance Risk Decile	---	1.05	~1.08	---
Recidivism Risk Decile	---	***1.21	***1.23	***1.24
Age	0.99	---	1.02	1.02
Number Prior Prison Incarcerations	**1.12	---	1.06	~1.08
Paroled to Region III	***.41		***0.44	***0.43
Recidivism Risk of Principal Commitment Offense	1.02		1.01	---
African American	~1.52	---	1.45	---
Mexican	0.89	---	0.76	---
Latino	**2.33	---	*2.23	*2.09
Test Accuracy (*c*)	0.68	0.67	0.72	0.71
Likelihood Ratio Chi-Square	43.33	42.35	73.26	63.16

Note: ~: $p < .10$; *: $p < .05$; **: $p < .01$; ***: $p < .001$; two-tailed tests.

Across the various models, it can be seen that three variables consistently emerge as strong, independent predictors of recidivism: the COMPAS Recidivism Risk Scale, Paroled to Region III, and Latino. While neither of the first two models achieved an overall accuracy score of at least .70, the combined models had a test accuracy of .71 and .72.

Using a proportional hazard model, the next analysis (Table 4) sought to predict the length of time parolees successfully remained out of prison after their index release. This analysis allowed us to test for the effects of these various predictor variables on parolees' *survival*, that is, time to first re-incarceration. Consistent with the logistic regression models summarized above, this model included standard predictors available from OBIS and all four COMPAS risk scales in various combinations. Because these analyses allowed us to include the full sample (rather than limiting the sample to those who had been released for at least one year), we were able to include females.

Table 4. Parameter Estimates from Survival Analysis Modeling Number of Days until Reincarceration as a Function of Parolee Characteristics and COMPAS Risk Score Deciles (N=25,110).

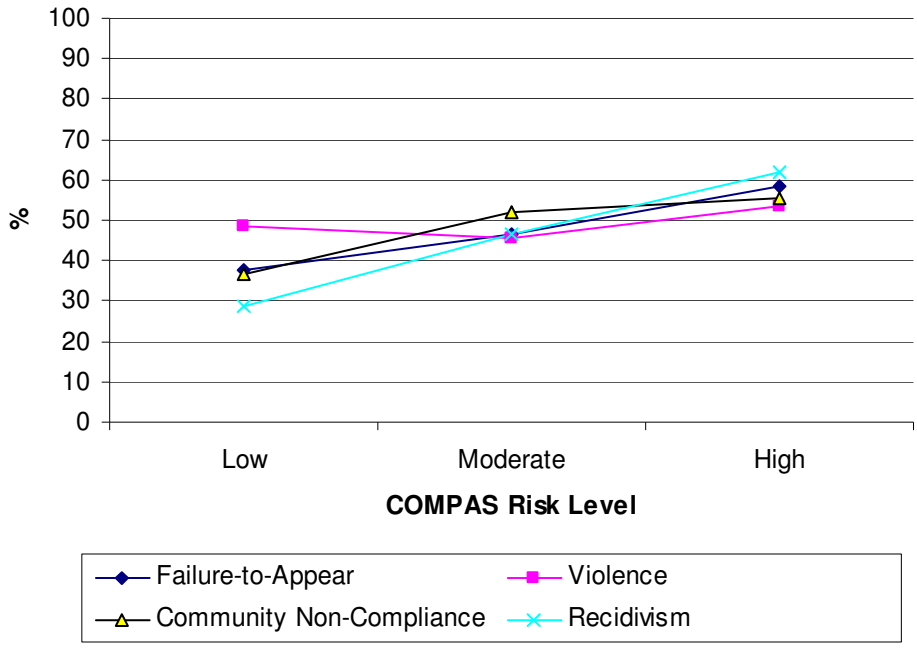
Predictor	Model 1	Model 2	Model 3	Model 4
Failure-to-Appear Risk Decile	---	***-0.04	***-0.03	***-0.03
Violence Risk Decile	---	***-0.02	**-0.01	**-0.01
Community Non-Compliance Risk Decile	---	***-0.02	***-0.02	**-0.02
Recidivism Risk Decile	---	***-0.09	***-0.07	***-0.07
Re-Release to Parole following Return to Custody	***-0.41	---	***-0.38	***-0.38
Female	***0.51	---	***0.54	***0.54
Age at Parole Release	***0.02	---	***0.01	***0.01
Number Prior Prison Incarcerations	***-0.07	---	***-0.05	***-0.05
Paroled to Region III	***0.63		***0.63	***0.62
Recidivism Risk of Principal Commitment Offense	***-0.02		***-0.01	***-0.01
African American	***-0.14	---	***-0.11	**-0.09
Mexican	***0.22	---	***0.21	***0.23
Latino	*-0.07	---	-0.05	---
Model Log-Likelihood	-15,916.55	-16,168.16	-15,714.59	-15715.61
Log-Likelihood Improvement over Intercept-Only Model	628.06	376.45	830.02	829.00

Note: *: $p < .05$; **: $p < .01$; ***: $p < .001$; two-tailed tests.

It is important to keep in mind that, with samples this large, even minor associations can be statistical significant. The parameter estimates shown in Table 4 demonstrate the magnitude and the direction of the relationships between the predictor variables and the *survival* following release. This table is included to show that, using a much larger sample, the COMPAS risk scales are significant predictors of the number of days in which an offender remains in the community after being released from prison. It also shows that combining the COMPAS risk scales with other variables from OBIS (e.g., region of parole) can strengthen the overall prediction models.

Figure 5 graphically illustrates the relationships between COMPAS risk levels (based on current cut-points) and being returned to custody with 12 months of release. As would be expected, the recidivism risk scale shows the strongest relationship with actual recidivism. The relationships between the other three scales and their specific outcomes (i.e., using the violence risk scale to predict return to custody for a violent offense) will not be testable until the outcome sample is substantially larger.

Figure 5: Percentage of Parolees Returned to Prison within 365 Days of Release by COMPAS Risk Estimates (N = 515)



Relationship between First-Order Scale Deciles and Recidivism

The final set of analyses for this report were conducted to ascertain the relationships between the first-order scales of the COMPAS (e.g., criminal involvement, substance abuse, etc.) and three types of criminal outcomes: return to custody, technical parole violation, and non-technical parole violation over the 12-month period following release. As shown in Table 5, the strongest relationships were found between the first-order scales and general return to custody, with 12 of the 18 scales showing a significant, positive relationship. Nine of the scales predicted returns for a non-technical parole violation. Only three of the first-order scales predicted technical parole violations.

Table 5. Spearman Rank-Order Correlations among COMPAS First-Order Scale Deciles and Three Indicators of Recidivism (All Parolees with One Year Observation Period, N=515)

	Returned to Custody within 365 Days		Technical Parole Violation within 365 Days		Non-Technical Parole Violation within 365 Days	
Scale	*r*	*p*	*R*	*p*	*r*	*P*
Criminal Involvement	0.16	***	0.14	**	0.09	*
History of Non-Compliance	0.12	**	0.09	*	0.10	*
History of Violence	0.07		0.10	*	0.07	
Current Violence	-0.03		-0.04		-0.02	
Criminal Associates/Peers	0.11	*	0.00		0.10	*
Substance Abuse	0.10	*	0.04		0.02	
Financial Problems/Poverty	0.06		-0.03		0.05	
Vocational Education Problems	0.17	***	0.05		0.12	**
Criminal Thinking	0.12	**	0.02		0.14	**
Family Criminality	0.01		0.03		0.00	
Social Environment Problems	0.16	***	0.02		0.10	*
Leisure and Recreation	0.10	*	0.00		0.11	*
Residential Instability	0.03		0.02		0.02	
Social Adjustment Problems	0.13	**	0.04		0.07	
Socialization Failure	0.17	***	0.01		0.15	***
Criminal Opportunity	0.18	***	-0.01		0.13	**
Criminal Personality	0.15	***	0.04		0.08	
Social Isolation	0.01		0.06		0.02	

Note: *: $p < .05$; ** $p < .01$; *** $p < .001$; two-tailed test.

V. Summary

The purpose of this evaluation is to assess the implementation of COMPAS, the validity of the *needs* scales, and the predictive validity of the *risk* scales. Regarding the implementation of the COMPAS, our site visits and staff interviews indicated that the C-File reviews take an average of 45 minutes, and the interviews typically last around 30-45 minutes. Among the perceived benefits of the COMPAS, according to the PSAs and PA IIs interviewed, were that the information from the offender needs assessment helps identify potential problems for supervision, such as homelessness, unemployment, treatment arrangements, and transitional housing. Staff also suggested that the COMPAS—and the resulting exit interview report—help the parolees' agent of record to quickly access the summary information in his first meeting with the parolee. PA IIs indicated that another strength of the COMPAS is its flexibility. Specifically, although they acknowledged the advantage

of having empirical backing of their decisions regarding supervision and referrals, they also appreciated that the COMPAS scores can be overridden by the parolees' AORs and/or unit supervisors.

Perceived problems with the COMPAS mainly centered around the need to reword several of the questions (particularly those involving double negatives), and the lack of uniformity in how the interviews are conducted. The PSAs also reported that the list of available community-based programs and services should be updated more regularly. No data were available regarding the extent to which the parolees availed themselves of the services on the referral lists provided to them as part of the exit interview.

Regarding the predictive efficacy of the COMPAS risk scales, this report included a summary of the initial preliminary findings (submitted to CDCR in September of 2007), which indicated that all four of the major COMPAS risk scales (violence, recidivism, failure to appear, and community non-compliance) were significantly correlated with at least one of three outcome measures (being returned to custody, having a technical parole violation, and having a non-technical parole violation). However, the area-under-the-curve (AUC) value for the recidivism scale was .67, and .61 for the community non-compliance scale.

To explore possible ways by which the existing COMPAS prediction models might be improved, a series of logistic regressions were conducted consisting of standard predictors available from OBIS and all four COMPAS risk scales, and various combinations of these sets of predictors. Across the various models, three variables consistently emerged as significant predictors of recidivism: the COMPAS Recidivism Risk Scale, being paroled to Region III, and being Latino. While neither of the first two models achieved an overall accuracy score of at least .70, the combined models had a test accuracy of .71-.72. These findings were replicated using a proportional hazard model, which allowed the inclusion of 25,110 parolees in the outcome validation sample.

The primary goals of the second year of this evaluation will be to (1) assess the reliability and validity of the needs scales by comparing scores from these scales with scores obtained on similar, more established scales, (2) conduct an updated assessment of the predictive validity of the COMPAS risk scales on a larger sample of parolees, and (3) incorporate the latest development of the COMPAS instrument by Northpointe into our analysis and continue to explore ways to improve its predictive validity. It is anticipated that findings related to all three of these goals will be available by the time of the second quarterly report to be submitted to CDCR in June of 2008.

APPENDIX

Table 1. *Spearman Rank-Order Correlations among Major COMPAS Risk Scales and Return-to-Custody for Any Reason within 365 Days of Release (All Parolees with One-Year Observation Period, N=515)*

	Returned To Custody within 365 Days	Technical Parole Violation within 365 Days	Non-Technical Parole Violation within 365 Days	Failure-to-Appear Risk Score Decile	Violence Risk Score Decile	Community Non-Compliance Risk Score Decile	Recidivism Risk Score Decile
Returned to Custody within 365 Days	1.00						
Technical Parole Violation within 365 Days	0.29	1.00					
Non-Technical Parole Violation within 365 Days	0.62	-0.06	1.00				
Failure-to-Appear Risk Score Decile	0.16	0.11	0.09	1.00			
Violence Risk Score Decile	0.05	0.09	0.03	0.12	1.00		
Community Non-Compliance Risk Score Decile	0.17	-0.01	0.17	0.55	0.03	1.00	
Recidivism Risk Score Decile	0.27	0.07	0.22	0.23	0.08	0.34	1.00
Mean	0.49	0.10	0.48	6.43	5.62	6.38	6.61
Standard Deviation	0.50	0.30	0.50	2.84	3.07	2.83	2.67

Note: Cutoffs for two-tailed inference tests: $r > .08$, $p < .05$; $r > .11$, $p < .01$; $r > .14$, $p < .001$.