RESEARCH SYNTHESIS
RELIABILITY AND VALIDITY OF COMPAS

Tim Brennan, Ph.D.
Bill Dieterich Ph.D.
Beate Ehret, Ph.D.

September 2007

Table of Contents

Executive Summary

Over the past several years a substantial body of research on the reliability and validity of the COMPAS assessment system has accumulated across various corrections jurisdictions nationwide. This report organizes and synthesizes the findings from a number of disparate psychometric and validation studies. Specifically, this report provides a *consistency* validation by evaluating the reliability and validity of the COMPAS scales across those different studies. The instrument is theoretically and empirically sound, with psychometric properties generally in the excellent or satisfactory range. Available findings indicate high predictive power of the COMPAS *Recidivism Risk* scale with AUC[1] scores above .7 for male and female correctional populations. This integrated report demonstrates the emerging consistency and generalizability regarding these findings.

The following psychometric properties of the COMPAS scales were tested across different study sites:

- **Scale Reliability.** The majority of the COMPAS scales have **high internal consistency** as evaluated using Cronbach's alpha with the balance providing measures of scale reliability in the satisfactory range. These findings generally hold across the sites, indicating that COMPAS is administered in a consistent and reliable manner across the different jurisdictions included in this comparison. Gender differences in alphas were only detected for very few of the scales used to rate the subjects. For all sites, the vast majority of the COMPAS scales were found to be **equally reliable for males and females.**

- **Internal Factorial Validity**. Due to our scale development strategy, **factorial validity of the COMPAS scales is ensured.** Each COMPAS subscale was developed as a unidimensional linear function. In general, items were selected to be both theoretically related to a certain construct and to load heavily on the same "component" or factor applying rotated principal component analysis. Variations of the factorial validity across different sites were found to be only minor and are not discussed in this report. The *Criminal Involvement* scale is presented as one example for the generally high internal validity and dimensionality of the scales.

- **Criterion/Concurrent Validity.** Observed relationships between the COMPAS subscales and criminal history indicators data provide **good evidence for criterion-related, concurrent validity** of the scales. The general **consistency of correlational patterns with criminal history indicators across different sites and datasets,** adds additional

[1] AUC scores are measures for the strength of association, for example between a scale and an outcome measure. AUC refers to the "area under the curve." This curve is created by plotting the true and false positive rates for different cut-off scores of a scale. The steeper the curve, the larger the area under the curve, and the better the scale's classification.

empirical support to this conclusion. Interestingly, no major differences regarding these patterns were found between offenders on probation or parole. Supplementary receiver operator characteristics analyses evaluate the *accuracy* of the scales in discriminating low and high-risk cases and confirm these findings with AUCs ranging between .71 and .82. These high AUC scores may not, however, support further generalization (e.g., to predictive validity), since realistic outcome measures (collected independently from the scale scores) were unavailable for this particular analysis.

- **Predictive Validity.** The predictive utility of the COMPAS risk scales is examined with survival models using different types of outcome measures. For a community corrections center in Ohio it is demonstrated that **the *Recidivism, FTA,* and *Violence* risk scales are strong predictors for the criterion of *program failure***. Inmates participating in a community service program, who scored high on the *Recidivism Risk* scale for example, have a risk or "hazard" of program failure that is three times higher than the failure hazard of inmates scoring low risk.

  More importantly, the COMPAS risk prediction model, now called the *Recidivism Risk* scale is tested using different *recidivism* outcome measures such as *New Crime, Any Offense*, *Person Offense*, and *Felony Offense*s. These survival analyses demonstrate a **high predictive power of the COMPAS *Recidivism Risk* scale.** An AUC of .72 as originally found in the scale construction sample of New York probationers predicting New Crime(s) is cross validated in a new sample with **AUC's ranging between .69 for *Any Offense* and .71 for *Offense(s) Against Persons*.** Since the AUCs in the validation sample are only slightly smaller than those in the scale construction sample, the **COMPAS *Recidivism Risk* prediction model can be considered robust. This also indicates that the risk model can be generalized to other offender populations with little loss of predictive accuracy.**

  Moreover, the predictive utility of the COMPAS *Recidivism* and *Violence-R Risk* scales is tested for parolees in a California sample using cause-specific Cox proportional hazards models. **Significant differences between offenders on the different risk levels of the *Recidivism* and *Violence-R Risk* scales were found regarding the risk of parole failure.** For example, inmates classified as high-risk on the *Recidivism Risk* scale have a failure hazard that is about 4 times higher than the hazard for inmates classified as low risk. Parolees on the high risk level of the *COMPAS Matrix-R* have a 4.7 times higher failure hazard in comparison to their counterparts on the low risk level. This study yielded rather low but nevertheless encouraging AUCs of .66 for both risk scales. Although the sample is very large, with more than 20,000 parolees, the follow-up is not mature yet with an average length of about 6 months. Therefore, the findings could change as additional follow-up time accrues.

- *Construct Validity:* Recent research provides evidence that the **COMPAS risk scales and their components are closely related to the construct of interest—recidivism.** The main risk scales (Recidivism and Risk Matrix) and their components (Criminal Involvement, Vocational/Educational Problems, and Drug Problems) consistently emerge in various meta analyses as major risk factors for recidivism (Lipsey & Derzon 1998, Gendreau, Little, & Goggin, 1996; Andrews & Bonta, 1994).

Given that instrument validation is an ongoing process, numerous further tests and models may be applied to further examine the concurrent and predictive validity of the COMPAS risk scales. The present overview, however, summarizes encouraging evidence that the main COMPAS risk scales, *Recidivism Risk* and the *Risk Matrix-R,* perform well in predicting recidivism in a variety of different criminal justice populations. Our next steps to further confirm these findings are additional outcome validation studies in larger samples and with longer follow-up periods. These analyses will be conducted at several different sites.

Abstract

Over the past several years a substantial body of research on the reliability and validity of the COMPAS assessment system has accumulated across various jurisdictions nationwide. This report organizes and synthesizes the findings from a number of disparate psychometric and validation reports that were conducted at different times for specific state systems, counties, or regional areas. In this respect, the report provides a *consistency* validation by evaluating the reliability and validity of the COMPAS scales across different studies. The instrument is theoretically and empirically sound, with psychometric properties generally in the excellent or satisfactory range. Available findings indicate the predictive power of COMPAS with AUC scores above .7 for male and female correctional populations. This integrated report demonstrates the emerging consistency and generalizability regarding these findings.

1. Introduction

The COMPAS system is designed to measure risk, need and mediating factors for recidivism as established by research (see, for example Lipsey & Derzon 1998, Gendreau, Goggin, & Little, 1996,). Indeed, many of the items entering into the COMPAS scales are demonstrated predictors of recidivism (e.g., Cottle, Lee, & Heilbrun 2001).

The COMPAS system calculates four different risk equation models specifically fitted for each of the following risk dimensions:

- Violence
- Recidivism
- Failure to Appear
- Community Failure

Moreover, there are several COMPAS instruments designed according to the type of criminal justice population:

- COMPAS Core
- COMPAS Probation
- COMPAS Re-Entry

These instruments account for variations in risk factor profiles at different stages of the criminal justice system. As opposed to other risk assessment tools that claim to be valid across different populations, we acknowledge that using an assessment instrument for parole releasees, which has been originally developed for probationers, weakens its predictive power. Given that many of the risk/need domains and individual items of Core COMPAS or COMPAS Probation focus on the offender's life in the community, they would assess the current risk of prison releasees rather insufficiently and not very accurately (see also Austin, 2006). For these reasons the COMPAS Re-Entry instrument was developed specifically for parole populations

and includes additional scales such as *Early Onset* (of criminal offending), occurrences of *Prison Misconduct,* or the risk of *Housing Problems* upon release. This report focuses on Core COMPAS.

2. Reliability

Northpointe has conducted several studies providing specific tests of psychometric reliability and factorial validity of all COMPAS scales. This section provides an overview of reliability findings across different regions, and across gender. The internal scale structure and factorial validity is illustrated using the *Criminal Involvement* scale. For more detailed descriptions of all COMPAS scales please see COMPAS Technical Manual, 2006.

2.1 Regional findings regarding Reliability

For a scale to be useful it must be reliable. Reliability is consistent measurement. Generally, if the items in a scale are highly correlated (internally consistent), then the summated scale will be reliable. Cronbach's alpha provides a convenient measure of the reliability of a scale. By convention, alpha coefficients of .70 or higher indicate good reliability.

Alpha coefficients for the Core COMPAS scales across various study sites are listed in Table 2.1.1 The shaded cells indicate alphas *below* the .70 threshold for good reliability (alphas of .69 are considered reasonably close to that threshold and still satisfactory—therefore these cells are also not shaded). The Number of cases are listed by site in Table 2.1.1. Table counts vary by scale due to missing values.[2] Alpha coefficients are not computed for the *Risk of Technical Violence* and *Risk of Recidivism* scales because these scales are linear equations.

The majority of the COMPAS subscales listed in Table 2.1.1 have alphas above .70. In most cases, these findings hold across sites. Some scales consistently display alphas below the .70 threshold across sites. These are the *History of Non-Compliance, Current Violence, Family Criminality,* and *Social Adjustment* scale. The majority of the scales in COMPAS have high internal consistency with the balance providing satisfactory measures of internal consistency. These findings indicate COMPAS is administered in a consistent and reliable manner at all sites included in this comparison.

[2] For a given scale, if the values for any of the items were missing, the value for the scale was considered missing.

Table 2.1.1: COMPAS Scale Reliability

COMPAS Scales[a]	Scale Reliability as indicated by Cronbach's Alpha					
	Probation		Combined[c]	Parole		
	Dallas County Courts' PSI[b].	San Bernardino Probation Assessments	Wyoming Community/ Incarcerated	MI Pre-Release Assessments	CA Pre-Release Assessments	GA Pre-Release Assessments
Number of cases	n= 1,170	n =1,534	n =1,065	n =1,071	n =1,077	n=3,905
Criminal Involvement	0.89	0.90	0.87	0.85	0.79	0.83
History of Non-Compliance	0.56	0.62	0.68	0.66	0.57	0.56
History of Violence	0.70	0.72	0.68	0.66	0.71	0.63
Current Violence	0.67	0.62	0.64	0.62	0.67	0.66
Criminal Associates/Peers	0.71	0.76	0.76	0.74	0.80	0.70
Substance Abuse	0.78	0.79	0.70	0.78	0.74	0.76
Financial Problems/Poverty	0.72	0.71	0.75	0.77	0.70	0.70
Vocational/ Educational Problems	0.69	0.68	0.65	0.69	0.67	0.67
Family Criminality	0.63	0.63	0.66	0.63	0.63	0.59
Social Environment/ Neighborhood	0.82	0.80	0.77	0.87	0.81	0.80
Leisure/ Boredom	0.80	0.79	0.84	0.86	0.82	0.80
Residential Instability	0.63	0.68	0.69	0.68	0.71	0.65
Social Adjustment	0.60	0.58	0.61	0.59	0.53	0.52
Juvenile Socialization Problems	0.70	0.70	0.71	0.68	0.68	0.65
Criminal Opportunity	0.66	0.63	0.68	0.71	0.68	0.63
Social Isolation	0.79	0.80	0.84	0.84	0.78	0.77
Criminal Attitudes/ Cognitions	0.79	0.82	0.82	0.77	0.76	0.78
Criminal Personality	0.72	0.75	0.76	0.75	0.68	0.67
Risk of Failure to Appear (FTA)	0.72	0.76	0.76	0.72	0.70	0.66
Risk of Violence	0.74	0.72	0.71	0.69	0.72	0.71

a Alphas are not available for the *Risk of Technical Violence* and *Risk of Recidivism* scales because these scales are linear equations.
b Pre-Sentencing Investigation **c** Prison/parole/probation/pretrial/jail/community corrections

2.2 Reliability across Gender breakdowns

To assess if the COMPAS scales are equally reliable for males and females, additional alphas are provided by gender for probation (Table 2.2.1) and parole (Table 2.2.2) samples. Confidence intervals of the difference in alphas between females and males were also computed to evaluate for significance of such gender differences.

For all sites, the vast majority of the COMPAS scales were found to be equally reliable for both genders. Few scales revealed gender differences in alphas. These were the *History of Non-Compliance* and *Social Environment/Neighborhood* scales in the Dallas County data, the *History of Violence* and *Risk of Violence* (which includes items from the *History of Violence* scale) scales in San Bernardino, and the *Leisure/Boredom*, *Residential Instability* and *Social Isolation* scales in Wyoming's combined data. Within the California parole sample, we found differences between males and females for the *Criminal Personality* and *Risk of Violence* scales. No gender differences regarding scale reliability were found in the Michigan parole sample. Overall, the scale differences in alphas across genders are inconsistent from site to site.

Table 2.2.1: COMPAS Scale Reliability–Males and Females in Probation and Combined Samples

COMPAS Scales	Scale Reliability as indicated by Cronbach's Alpha					
	Dallas County Courts' PSI[b]		San Bernardino Probation Assessments		Wyoming Community/ Incarcerated	
	Female	Male	Female	Male	Female	Male
Criminal Involvement	0.90	0.89	0.91	0.90	0.88	0.85
History of Non-Compliance	0.40*	0.57*	0.62	0.62	0.65	0.61
History of Violence	0.72	0.69	0.57*	0.72*	0.60	0.67
Current Violence	0.69	0.67	0.64	0.61	0.68	0.70
Criminal Associates/Peers	0.70	0.70	0.75	0.77	0.77	0.75
Substance Abuse	0.79	0.78	0.78	0.79	0.69	0.70
Financial Problems/Poverty	0.72	0.72	0.68	0.72	0.75	0.74
Vocational/Educational Problems	0.70	0.69	0.65	0.69	0.67	0.66
Family Criminality	0.60	0.62	0.61	0.63	0.64	0.67
Social Environment/ Neighborhood	0.77*	0.83*	0.79	0.81	0.73	0.78
Leisure/ Boredom	0.82	0.79	0.79	0.79	0.88*	0.83*
Residential Instability	0.62	0.64	0.67	0.69	0.63*	0.71*
Social Adjustment	0.64	0.59	0.56	0.60	0.62	0.62
Juvenile Socialization Problems	0.69	0.71	0.66	0.72	0.73	0.71
Criminal Opportunity	0.68	0.65	0.62	0.63	0.71	0.68
Social Isolation	0.81	0.78	0.81	0.80	0.87*	0.83*
Criminal Attitudes/Cognitions	0.76	0.79	0.79	0.82	0.80	0.82
Criminal Personality	0.72	0.71	0.72	0.77	0.73	0.76
Risk of Failure to Appear (FTA)	0.72	0.72	0.77	0.77	0.76	0.77
Risk of Violence	0.72	0.74	0.65*	0.72*	0.67	0.70

* The difference in alphas between males and females is considered significant if the 95% confidence interval does not include zero.
b Pre-Sentencing Investigation **c** Prison/parole/probation/pretrial/jail/community corrections

Table 2.2.2: COMPAS Scale Reliability – Males and Females in **Parole** Samples

COMPAS Scales	Scale Reliability as indicated by Cronbach's Alpha			
	MI Pre-Release Assessments /w inmates		CA Pre-Release Assessments	
	Female	Male	Female	Male
Criminal Involvement	0.82	0.86	0.81	0.79
History of Non-Compliance	0.70	0.65	0.54	0.58
History of Violence	0.61	0.65	0.73	0.68
Current Violence	0.66	0.59	0.69	0.67
Criminal Associates/Peers	0.70	0.76	0.80	0.81
Substance Abuse	0.78	0.78	0.75	0.71
Financial Problems/Poverty	0.78	0.76	0.66	0.71
Vocational/Educational Problems	0.69	0.70	0.66	0.69
Family Criminality	0.64	0.61	0.59	0.62
Social Environment/ Neighborhood	0.86	0.87	0.83	0.81
Leisure/ Boredom	0.85	0.86	0.82	0.81
Residential Instability	0.70	0.67	0.71	0.71
Social Adjustment	0.60	0.59	0.54	0.53
Juvenile Socialization Problems	0.68	0.68	0.68	0.68
Criminal Opportunity	0.72	0.70	0.69	0.68
Social Isolation	0.84	0.84	0.80	0.78
Criminal Attitudes/Cognitions	0.74	0.77	0.72	0.76
Criminal Personality	0.75	0.76	0.73*	0.67*
Risk of Failure to Appear (FTA)	0.70	0.74	0.74	0.69
Risk of Violence	0.67	0.66	0.76*	0.69*

* The difference in alphas between males and females is considered significant if the 95% confidence interval does not include zero.

2.3 Internal Structural Validation: Factorial Validity

Each COMPAS subscale was developed as a unidimensional linear function. In general, items were selected to be both theoretically related to a certain construct and to load heavily on the same "component" or factor. The factor patterns are based on rotated principal component analyses for each scale.

The importance of each scale item is indicated by the size of the factor *loading*. A large loading by an item on the first component (i.e., 0.4 and higher) indicates that the item is an important contributor to the scale. Given that several items with strong loadings are included on each scale, large *eigenvalues* for the first factor, or component are achieved. A large eigenvalue (greater than 1) indicates that the factor accounts for a large amount of overall variance in the summated score, which, in turn, suggests that the scale is measuring a *single* underlying construct. Due to this scale development strategy, factorial validity of the COMPAS scales is ensured. A comparison of the factorial validity across different sites reveals only minor variations (for details see COMPAS Technical Manual).

To provide one example, the *Criminal Involvement* scale is presented. It consists of four items indicating the number of: previous jail times "n.jails," previous arrests, "n.prev.arrest," previous convictions, "n.prev.convict," and previous episodes on probation, "n.probations."

The factor loadings of the *Criminal Involvement* scale are shown in Table 2.3.1 and the Eigenvalues in Table 2.3.2. This scale is clearly unidimensional with a large eigenvalue of the first component (PC1) 3.05, accounting for 76.2% of the variance of the item set. The strong factor-loading pattern on the first component confirms the importance of all items in the scale.

Table 2.3.1: Principal Component Pattern: Criminal Involvement in Dallas County Probation.

Items	PC1	PC2	PC3	PC4
n.jails	0.495	0.488	0.653	−0.301
n.prev.arrest	0.527	0.215	−0.192	0.799
n.prev.conv	0.518	0.058	−0.676	−0.520
n.probations	0.456	−0.844	0.282	−0.006

Table 2.3.2: Eigenvalues and Variance Explained in Dallas County Probation Data.

	PC1	PC2	PC3	PC4
Eigenvalue	3.047	0.483	0.283	0.188
Standard Deviation	1.745	0.695	0.532	0.433
Proportion of Variance	0.762	0.121	0.071	0.047

3. Criterion/Concurrent Validity

As part of our standard psychometric reports, a section on criterion related "concurrent" validity is included. In addition to correlations with criminal history indicators, *Receiver Operator Characteristics* are provided below, evaluating the *accuracy* of the scales in discriminating low and high-risk cases.

3.1 Concurrent Validity: Correlations with Criminal History Indicators

A common approach to assessing the concurrent validity of a scale is to determine whether the scale scores are correlated with the *criterion* behavior of interest such as *Age-at-first Arrest*, or *Number of Prior Arrests*. Two tables are provided (Table 3.1.1 and 3.1.2) to examine concurrent validity separately for offenders on probation and parole. Shaded cells indicate significant correlations.

For this concurrent, correlational approach, it is important to make sure that the criterion variable is itself not included as an item on the scale. In this section, Spearman or Pearson (for the Michigan and Georgia samples) correlations of the COMPAS scales against selected criminal justice and criminological criterion variables (e.g. *Age-at-first Arrest, Number of Previous Arrests*, etc) are presented across different study sites.

Patterns of correlations emerge at the probation and parole sites compared in Tables 3.1.1 and 3.1.2. For example, *Age-at-first Arrest* correlates negatively with the personality scales *Criminal Personality* and *Criminal Attitudes*. These results comport with findings in developmental research indicating offenders with early onset are more likely to have high scores on similar types of personality measures and more serious and persistent criminal involvement (Moffitt, 1993). The observation also corresponds with the positive correlation of the *Criminal Personality* scale and *Total Number of Previous Arrests*. Similar to the above, offenders with earlier *Age-at-first Arrest* are more likely to have higher scores on scales measuring factors that have been identified as criminogenic in longitudinal criminal justice research studies. These scales include *Social Environment*, *Vocational/Educational Problems*, and *Family Criminality* (Farrington et al., 2001).

Another pattern evident for probation and parole populations (Tables 3.1.1 and 3.1.2) is defined by the correlations of *Previous Arrests* as well as *Probation/Parole Revocations* with the scales *Criminal Associates/Peers* and *Substance Use* which represent risk factors that are highly relevant for predicting criminal involvement as established by research (Stouthamer-Loeber et al., 2002; Cottle, Lee & Heilbrun, 2001).

Table 3.1.1: Spearman Correlations of COMPAS Sub-scales with selected Criminal History Indicators in *Probation* and *Combined* Samples

Probation	**Age-at-First**			**Previous Arrests**			**Probation Revocation**		
	Dallas	**San Bernar-dino**	**Wyoming** combined sample*	**Dallas**	**San Bernar-dino**	**Wyoming** combined sample*	**Dallas**	**San Bernar-dino**	**Wyoming** combined sample*
Scale	N=1,148	N=1,047	N=988	N=1,148	N=1,047	N=988	N=1,148	N=1,047	N=988
Criminal Associates	-0.26	-0.14	-0.27	0.24	0.32	0.22	0.19	0.25	0.17
Substance Abuse	-0.02	-0.04	-0.18	0.25	0.34	0.29	0.19	0.34	0.19
Financial Problems	-0.02	0.01	0.02	0.04	0.17	0.01	0.07	0.14	-0.03
Vocational Educational	-0.29	-0.15	-0.26	0.09	0.19	0.12	0.12	0.15	0.12
Family Criminality	-0.12	-0.09	-0.20	0.05	0.14	0.14	0.08	0.12	0.13
Social Environment	-0.17	-0.11	-0.18	0.07	0.17	0.14	0.06	0.19	0.08
Leisure/ Boredom	-0.10	-0.06	-0.24	0.03	0.18	0.10	0.06	0.17	0.11
Residential Instability	-0.03	-0.03	-0.10	0.05	0.04	0.14	0.02	0.07	0.10
Social Isolation	0.01	-0.03	-0.07	0.09	0.18	0.12	0.07	0.12	0.07
Criminal Attitudes	-0.13	-0.07	-0.17	0.01	0.10	0.06	0.01	0.12	0.06
Criminal Personality	-0.18	-0.11	-0.29	0.10	0.22	0.22	0.02	0.20	0.16

Note: With N=1148, a correlation of .058 is significant, with N=1047, a correlation of .06 is significant, and with N=988, a correlation of .063 is significant at $p < .05$ (2-tailed). * Including prison/parole/probation/pretrial/jail/community corrections cases.

While some of these generally robust patterns were not confirmed in the Michigan parole data (Table 3.1.2), overall, the observed relationships between the COMPAS subscales and criminal history indicators data provide evidence for criterion-related, concurrent validity of the scales. Their general consistency across different sites and datasets, adds additional empirical support to this conclusion. Interestingly, no major differences in correlational patterns with criminal history indicators were found between offenders on probation and parole. On the whole, significant correlations are less common amongst parolees, while the dominant patterns are the same between the two comparison groups.

Table 3.1.2: Spearman/Pearson Correlations of COMPAS Sub-scales with Selected Criminal History Indicators in *Parole* Samples.

Parole	**Age-at-First**			**Previous Arrests**			**Probation Revocation**		
	MI	**CA***	**GA**	**MI**	**CA***	**GA**	**MI**	**CA***	**GA**
Scale	N=920	N=785	N=3,809	N=920	N=785	N=3,809	N=920	N=785	N=3,809
Criminal Associates	−0.20	-0.20	−0.18	0.04	0.15	0.04	−0.20	0.07	0.04
Substance Abuse	−0.05	0.04	0.03	0.17	0.27	0.21	−0.05	0.19	0.15
Financial Problems	−0.01	-0.03	0.01	0.08	0.11	0.10	−0.01	0.12	0.14
Vocational Educational	−0.23	-0.24	−0.28	0.03	0.04	−0.03	−0.23	0.05	0.01
Family Criminality	−0.13	-0.09	−0.09	0.03	0.05	0.03	−0.13	-0.05	0.06
Social Environment	−0.17	-0.13	−0.10	−0.10	0.08	−0.06	−0.17	0.10	−0.03
Leisure/ Boredom	−0.04	-0.08	−0.07	0.06	0.08	0.01	−0.04	0.06	0.01
Residential Instability	0.01	-0.08	0.00	0.01	0.02	0.02	0.01	0.10	0.01
Social Isolation	−0.15	-0.02	−0.01	0.06	0.10	0.04	−0.15	0.04	0.05
Criminal Attitudes	0.03	-0.17	−0.14	0.09	0.06	−0.06	0.03	0.11	−0.01
Criminal Personality	−0.12	-0.19	−0.16	0.07	0.12	0.00	−0.12	0.07	0.03

With N=920 a correlation of .065 is significant, with N=785, a correlation of .07 is significant, and with N=3809, a correlation of .032 is significant at $p < .05$ (2-tailed). * For California Spearman correlations are presented, for Michigan and Georgia Pearson correlations are provided.

3.2 Concurrent Validity: Receiver Operator Characteristics (ROC)

For a New York Probation sample, concurrent validity was also evaluated by means of Receiver Operator Characteristics (ROC). This method can be applied to assess how accurately the risk scales discriminate between *Recidivists* and *Non-Recidivists*. ROC analysis produces a plot of the "sensitivity" of the scale against one minus the "specificity" of the scale, called the "ROC curve." Sensitivity is the percentage of *failures* correctly classified, also called the "true positive rate," and specificity is the percentage of *successes* correctly classified. One minus specificity is the false positive rate, in other words the percentage of offenders classified as Recidivists although they are not recidivating. Accordingly, the ROC curve is a plot of the true positive rate against the false positive rate for each cut-point of the scale. The Area Under the Curve, (AUC) gives a measure of the accuracy of the scale in discriminating between true and false positive rates. Larger areas under the curve indicate better accuracy. The AUC indicator can be interpreted as the probability that a randomly selected *Recidivist* will have a higher risk score

than a randomly selected *Non-Recidivist*. An AUC of .70 or above indicates a strong association, while measures between .60 and .70 indicate a moderate association according to recent research in the field of offender classification (Quinsey et al., 1998, Jones, 1996, Aos & Barnoski 2003).

For the ROC analysis of the scales *Risk of Community Non-Compliance, Risk of FTA,* and *Risk of Violence* the outcome variables were excluded from the scales and used to test *concurrent* validity without true follow-up data. For this purpose, these outcome measures were dichotomized so that 0 represents no failure and 1 represents 1 or more failures. As presented in Table 3.2.1, the AUC measures of all three scales are above the .70 threshold for a strong association, suggesting these scales have high concurrent validity. However, since more realistic outcome measures were unavailable, it is important to be cautious about generalizing these results on concurrent validity to predictive validity. The AUC values may be biased upward because of undesirable correlations between the scale items and the outcome variables selected for testing. Such correlations can be due to a measurement effect, if scale items and outcome variables used for the analysis were measured with the same instrument at the same point in time.

Table 3.2.1: COMPAS Concurrent Validity–Area under the Curve

COMPAS Scales	**Concurrent Validity as Indicated by AUC's**
	New York Probation
Sample Size	n= 393
Risk-Community Non-Compliance (CNC)	0.79 n.prob.rev*
Risk of Failure to Appear (FTA)	0.82 n.fta*
Risk of Violence	0.71 Violence**

* This Item was removed from the scale and dichotomized (0= no failure; 1=1 or more failures).

** This Item is a combination of items that were removed from the scale and indicates if the current offense involved violence at time of the assessment interview (0= no 1= violent).

4. Predictive Validity—Survival Analysis

Several agency or regional-specific studies have been conducted in which the basic COMPAS scales or the risk prediction scales (*Recidivism Risk, Violence Risk, FTA Risk,* and *Risk of Technical Violation*) have been evaluated for predictive validity. Techniques of survival analysis have been used to estimate "Proportional Hazard" models. These models were selected because they not only examine the *occurrence* of a recidivism event but also the *timing* of that event. This section will review the findings of these studies and offer conclusions regarding the generalizability of the predictive models that have been evaluated on different criminal offender populations (Probationers, Parolees, Inmates).

4.1 Community Corrections Ohio

A study in Ohio evaluated the utility of COMPAS risk scales in predicting community corrections *program failure*. The study sample was a subset of inmates (N=493) who were assessed with the COMPAS instrument from January 2002 through July 2005 and participated in a community corrections program that places inmates at government and non-profit agencies to perform community service.

Program failure was used as the outcome of interest and program success as the "competing" event. Although it may not be readily apparent, these outcomes represent competing risks. The analytic approach should adjust for the fact that subjects who experience the competing event of program success are no longer at risk of program failure, and vice versa. Therefore, a specialized proportional hazards survival model for competing risks was applied (Fine & Gray, 1999).

The relative failure probabilities of offenders on the high, medium and low risk levels of the COMPAS risk scales were compared using Hazard Ratios. Since the risk levels are coded as dummy variables, these Hazard Ratios can be interpreted as multipliers of the risk or "hazard" of program failure for a particular group (e.g. high risk offenders) compared to the reference group (here: low risk offenders).

Results from survival models as presented in Table 4.1.1 show that the *Recidivism, FTA,* and *Violence* risk scales are strong predictors for the criterion of program failure. Inmates scoring high on the *Recidivism Risk* scale for example, had a risk of program failure or "hazard" that is three times higher than the failure hazard of inmates scoring low risk. While these increased hazard rates of high compared to low risk offenders were not as large for the *Violence* (2.4 times higher) and *FTA* (1.9 times higher) risk scales, a significant difference was found for these scales as well. On the *Technical Violation* scale, medium and high risk offenders were not significantly different from offenders classified as low.

The findings demonstrate clearly that the COMPAS risk scales are valid tools for determining eligibility for one of the core community corrections programs at this corrections center in Ohio.

Table 4.1.1: Competing Hazards Model Regressing Program Failure on COMPAS Risk Scales: Hazard Ratios and p-values.

N=493		**Program Failure**	
Scales	**Risk Level**	**Hazard Ratio**	**p- value**
Technical	Medium	0.96	0.872
Violation	High	1.42	0.268
		Hazard Ratio	**p- value**
Risk of	Medium	1.39	0.182
FTA	High	1.94	0.008
		Hazard Ratio	**p- value**
Risk of	Medium	1.76	0.017
Violence	High	2.36	0.043
		Hazard Ratio	**p- value**
Risk of Recidivism	Medium	1.75	0.034
	High	3.00	0.000

Note. The reference category for the test of medium and high risk categories is low risk.

4.2 New York Probation—Scale Construction

A Study in a New York Probation sample examined the predictive validity of several recidivism risk scales. The outcome or *criterion* behavior of interest for this evaluation was recidivism or more specifically a *"New Crime."*

During 1999 COMPAS (Version 2) assessment data for offenders (n=513) was collected at correctional institutions of four sites in different counties in New York (Fulton, Monroe, Schenectady, and Suffolk). Subsequently, independent two-year follow-up data was made available for offenders in this COMPAS sample from the State Criminal Justice Database.

Many of the offenders in the original assessment sample were restricted from entering the community after their COMPAS screening date. For example, some were given prison time. For this reason, 103 cases with prison sentences were excluded from the sample. Seventeen cases with missing dispositions were also excluded. The remaining sample is composed of probation, jail, conditional discharge, and split sentence cases (n=393). These offenders were generally unconstrained from entering the community. In other words, they were "at risk" to recidivate. Although offenders with jail or split sentences were constrained for certain unknown periods of time, they were still considered "at risk" to recidivate, since these periods of detention were relatively short. This assumption was further confirmed by their relatively high recidivism rates during the two-year follow-up (e.g. jail 55%, split sentence 39%).

The follow-up data indicated, whether an offender was accused of committing a new crime and the date on which this alleged crime occurred. Based on this information the two outcomes variables "New Crime" and "Days until New Crime Occurred" were calculated. Predictive models of recidivism were developed with several different "candidate sets" of

variables using standard logistic regression. For the purpose of this report, a Proportional Hazard model based on techniques of survival analysis and a final logistic regression model with a similar final set of predictor variables are presented below.

The Cox Proportional Hazard model presented in Table 4.2.1 is based on a large candidate set of predictor variables, including all of the COMPAS basic scales and the three additional items *Age, Age-at First Arrest*, and *Arrest Rate*. This large set of variables was reduced using the Akaike Information Criterion (AIC) model selection procedure. AIC is a stepwise model selection procedure, which reduces the number of predictor variables, similar to stepwise regression procedures, according to the strength of their effect (Venables and Ripley, 1999).

The COMPAS scales aim at measuring risk, need, and mediating factors of recidivism, hence they were all included in this model. For a more detailed description of the scales included in the model see Appendix A. Moreover, previous research has shown that the age of an offender at the time of the arrest is an important predictor of recidivism. The numbers of arrests over the course of a criminal career are a good indicator of prior criminal involvement and also predict recidivism. Therefore, these variables were included in the model. The variable *Age-at First Arrest* represents the age at which the offender was first arrested. The variable *Arrest Rate* stands for the average number of arrests per year experienced by the offender prior to the assessment date.

The effects of covariates on the risk or "hazard" of recidivism can easily be interpreted based on the antilogarithm of their estimated coefficient "exp (coeff)," also called the "alpha effect" (Tuma and Hannan 1984). This alpha effect indicates the percentage change in the estimated hazard rate if the value of the covariate is increased by one unit (change of hazard rate=(exp(α) –1) * 100%). It takes the value 1 when the covariate has no effect, it is smaller than 1 when the effect is negative, and it is greater than one when the effect is positive. For example, the alpha effect of the *Criminal Involvement* scale (see Table 4.2.1) of 1.276 indicates a 27.6% increase of the hazard to recidivate for each increase of the scale's score by one unit. A unit on this scale stands for a prior instance of arrest, probation, conviction, or jail time.

The AIC model selection procedure was applied after regressing the variable "Days until New Crime Occurred" onto the predictors from the candidate set. "*New Crime*" served as the "censoring" variable, which refers to the "event" that constitutes the end of the timeline for this survival model.[3] It is important to stress that the accuracy of "*Days until New Crime Occurred"* can be questioned, especially for the offenders in the "jail" category, who were constrained for unknown periods of time.

The findings provide reasonable evidence for the predictive utility of several of the COMPAS scales. Table 4.2.1 lists the significant predictor variables that remained in the model after the AIC selection procedure.

[3] For an overview of survival analysis see Hosmer and Lemeshow, (1998).

Table 4.2.1: Cox Proportional Hazard Model Regressing Recidivism on COMPAS Risk Scales: Regression Coefficients and p-values.

	Recidivism		
PREDICTORS	**Coeff**	**EXP(Coeff)**	**P >|z|**
Criminal Involvement	0.244	1.276	1.4e-02
Financial Problems	0.331	1.392	8.5e-02
Vocational/ Educational Problems	0.427	1.532	1.3e-02
Use of Leisure Time	-0.364	0.695	2.4e-02
Residential Instability	0.240	1.271	7.0e-02
Social Adjustment	-0.475	0.622	4.3e-02
Criminal Attitudes	-0.275	0.774	1.1e-01
Criminal Personality	0.401	1.494	1.5e-02
Age at 1st Arrest	-1.414	0.243	1.7e-04
Arrest Rate	0.611	1.842	3.6e-05

The higher offenders score on the *Criminal Involvement, Financial Problems, Vocational/Educational Problems, Residential Instability,* and *Criminal Personality* scale, the higher their risk to recidivate. As often occurs when variables are collinear (closely correlated), some of the signs of the parameters are in unexpected directions such as for *the Use of Leisure Time, Social Adjustment,* and *Criminal Attitudes* scales.

The estimated alpha effects indicate changes of the hazard for each increase by one unit (scale score) of the predictor variables. The higher an actual offender's score, the higher their cumulative increase of the hazard per unit compared to low scores. Hence, COMPAS scales with positive effects represent a substantially increased risk to recidivate for offenders with high scores.

As indicated by the negative coefficient of *Age-at First Arrest*, an early onset of criminal involvement, predicts an increased risk to recidivate. The average number of arrests per year *(Arrest Rate)* can be seen as an indicator for a criminal history prior to the COMPAS assessment. With every additional prior arrest per year, the hazard rate for recidivism increases by 84%. According to the p-values, all effects are significantly different than zero.

A comparable logistic regression model predicting recidivism (derived from a reduced initial set of variables entering the AIC model selection procedure) includes an almost identical final set of predictor variables as shown in Table 4.2.2. This final, simpler model was selected and employed as the risk equation of the *COMPAS Recidivism Risk* scale. All effects are in the

expected direction. For a more detailed description of the predictor variable selection and model fitting process for the *Recidivism Risk* scale please see the COMPAS Technical Manual and Psychometrics Report.

The results of the logistic regression (Table 4.2.2) are very similar to the Hazard model indicating an increased risk of recidivism the higher offenders score on the *Criminal Involvement*, *Vocational/ Educational Problems* or *Drug Problems* scale. Also, the higher their average numbers of prior arrests per year *(Arrest Rate)* the more likely offenders are to commit a new crime. In this model, *Age-at First Arrest* and *Current Age* are included, both displaying effects in the expected direction—the older offenders were at their first arrest or at the time of the assessment, the lower their risk of recidivism.

As opposed to the Hazard model, the final set of the logistic regression includes an additional *Drug Problems* scale. This scale summarizes indicators whether the offender has had problems with drugs from the offense history, current charges, and the current arrest, in combination with information regarding a history of drug treatment or an expressed current drug treatment need (for more details see Appendix A).

Table 4.2.2: Pearson Correlations with New Crime (*r*) and Logistic Regression Model Predicting New Crime (Penalized Odds Ratios)

PREDICTORS	r	Odds Ratio (Penalized)
Intercept		-0.35672
Criminal Involvement	0.20	0.26443
Vocational/ Educational Problems	0.22	0.28935
Age	-0.18	-0.04567
Age at 1st Arrest	-0.28	-1.30161
Arrest Rate	0.29	0.46997
Drug Problems	0.22	0.54287

The sample selection may have caused a sample bias since offenders with prison sentences who are more likely to be serious offenders were excluded. In addition, some offenders were not “at risk” to recidivate but constrained for unknown periods of time (jail or split sentences). However, both these sample selection issues make it less likely to detect differences in recidivism between low and high risk offenders because they limit the variance of risk in the sample. Hence the sample bias is *not in favor* of the tested hypothesis, that the scales have the potential to discriminate between low and high risk offenders. Based on this assumption, this study can be seen as a rather “conservative’ test of the predictive validity of the COMPAS risk scales and the findings are convincing.

Receiver Operator Analysis of Recidivism Risk Scale

The logistic regression based on a reduced set of covariates as presented in Table 4.2.2, lead to the development of the *Recidivism Risk Scale* in the New York Probation sample. The components of the *Recidivism Risk* Scale are the *Criminal Involvement, Vocational Educational Problems,* and *Drug Problems* scales, together with the offender's age at the time of the assessment, at the time of the first arrest, and the average arrest rate per year prior to the assessment.

Chart 4.2.1 below illustrates the relationship between the *Recidivism Risk* scale and the probability of recidivism (committing a new crime) for the New York Probation sample (described above). The risk of recidivism clearly increases across quartiles of the scale.

The *Risk of Recidivism* scale was further validated applying ROC analysis in the New York Probation data, based on the outcome measure "New Crime"; hence the ROC analysis in Table 4.2.3 below evaluates the *predictive* validity of this scale. The values in the column labeled AUC* are estimates of AUC that have been corrected for bias with bootstrap validation. The correction accounts for the validation being performed in the scale construction sample. The scale may perform less well in a new sample. The findings are positive, with an AUC of .74 and a corrected AUC of .72, indicating high predictive validity of the *Risk of Recidivism* scale. The AUC values are similar in magnitude to those obtained by other researchers in the field (Cottle, Lee, & Heilbrun, 2001; Gran, Belfrage, & Tengstrom, 2000; Quinsey et al., 1998); and higher than for comparable fourth generation risk assessment instruments (Barnoski & Aos 2003).

Chart 4.2.1: Recidivism Risk Scale and Recidivism Rate

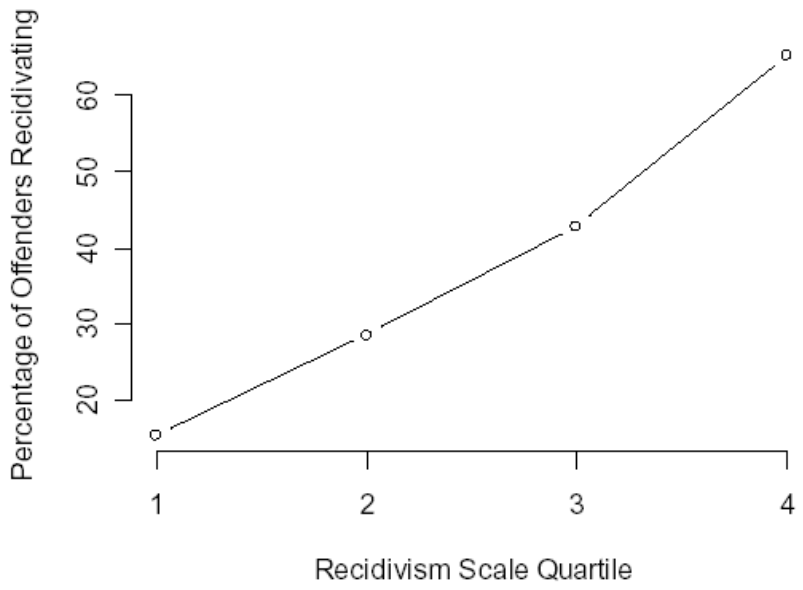


Table 4.2.3: COMPAS Predictive Validity–Area under the Curve

COMPAS Scale	AUC		AUC*	
Risk of Recidivism	0.74 New (within 2 yrs.)	Arrest	0.72* New (within 2 yrs.)	Arrest

*Corrected for bias by bootstrap validation.

4.3 New York Probation—Cross Validation

A second recidivism study was conducted to determine how predictive the COMPAS *Recidivism Risk* scale is in a different "validation" sample, a step that is also called "cross validation." The risk prediction model, as derived from the scale construction sample described above (see Table 4.2.2), is applied to 2,328 offenders in New York that were assessed with COMPAS at the point of Pre-Sentencing Investigation (PSI) or probation intake. The data were collected in 19 counties that administer the complete COMPAS assessment. The assessments were conducted from January 2000 through December 2004.

The COMPAS assessment data were matched with Computerized Criminal History (CHH) records by the New York State Division of Criminal Justice Services. Multiple-record survival datasets were constructed using assessment dates and event dates in the CCH data including crime dates, arrest dates, dispositions, disposition dates, sentence type, and sentence length. Three types of recidivism were examined including *Any Offense*, *Person Offense*[4], and *Felony Offense*. Separate Cox regression models were fitted for each type of recidivism. These models do not just predict failure but the time until a recidivism event occurs starting from the time of the assessment. Moreover, these models control for periods of incarceration during the follow-up time by removing the subject from the risk sets during such gaps, since they are not "at risk" of recidivism while incarcerated.

Results of receiver operator characteristics analysis (AUCs) for the COMPAS *Recidivism Risk* Cox regression model predicting any offense, offenses against persons, and felony offenses are presented in Table 4.3.1 below. All of the AUCs are indicating a moderate to high predictive validity. It should also be pointed out that the AUCs in the validation sample are only slightly smaller than those in the scale construction sample. Therefore the COMPAS *Recidivism Risk* prediction model can be considered robust. This also indicates that the risk model can be generalized to other offender populations with little loss of predictive accuracy.

[4] The recidivism type "person offense" is defined as a finger printable arrest involving a charge for any CUCR code for murder, forcible rape, robbery, aggravated assault, simple assault, burglary (with weapon or occupied dwelling), dangerous weapons, sex offenses, extortion, arson or kidnapping. This category includes misdemeanor and felony offenses.

Table 4.3.1: COMPAS Predictive Validity–Area under the Curve in Validation Sample

	AUCs total sample (*n*=2,328)			**AUCs** women (*n*=449)			**AUCs** men (*n*=1,879)		
	any	person	felony	any	person	felony	any	person	felony
Risk of Recidivism*	.68	.71	.70	.65	.76	.66	.68	.70	.71

* The Recidivism Risk scale is coded as deciles in this model.

4.4 California Parole

An outcomes study in a large sample (n=20,898) of parolees in the California Department of Corrections and Rehabilitation (CDCR) examined the predictive validity of the COMPAS risk scales (Brennan & Dieterich, 2007). The parolees were released onto parole during the period March 2006 through June 2007. In this study the *COMPAS Risk Matrix-R* and its two components, the *Violence-R Risk* scale and *Recidivism Risk* scale, were evaluated for their utility in predicting parole failure. Parole failure is defined as a return to prison (RTP) for a nontechnical parole violation.

Since offenders can return to prison for other reasons such as technical violations, while others can be discharged from parole, these events were modeled as "competing risks." This approach adjusts for the fact that subjects who experience these competing events are no longer at risk of returning to prison for a nontechnical violation. To take this into account, a specialized proportional hazards survival model for competing risks was applied (Fine & Gray, 1999).

The *Risk Matrix-R* consists of two risk dimensions; the first is defined by the *Risk of Violence-R* decile score, and the second dimension by the *Recidivism Risk* decile score. The *Risk Matrix-R* as evaluated in this study differs from the original Core COMPAS *Risk Matrix* in that it includes the *Violence-R* scale, an adaptation of the *Violence* scale specifically for parole populations. For the purpose of this synthesis of validation findings, the evaluations of the *Recidivism Risk* scale and the *Risk Matrix-R* are presented.

As an illustration of the basic logic of survival techniques and the resulting model, the charts below depict failure rates over time by risk level. This is done by plotting the predicted values from a regression of the cumulative incidences of RTP on the covariate (i.e., the levels of the Recidivism Risk Scale or of the *Risk Matrix-R* (Fine & Gray, 1999). Chart 4.4.1 shows a plot of the fitted failure probabilities for the three levels of *Recidivism Risk*, and Chart 4.4.2 does so for the *Risk Matrix-R*. As depicted for the *Recidivism Risk* scale (Chart 4.4.1), the three risk levels display substantial differences in their failure dynamics. For example, offenders on the high-risk level reach a failure probability rate of almost 10% within 100 days upon release compared to much lower rates of the two comparison groups at medium or low risk. While the chart for the *Risk Matrix-R (*Chart 4.4.2) is interpreted in the same manner, four risk levels are displayed here. Offenders on all four risk levels reveal major differences in their failure dynamics over time.

Chart 4.4.1: Predicted Cumulative Incidence of Nontechnical Violation Returns to Prison (RTP) for the Risk Levels of the Recidivism Scale.

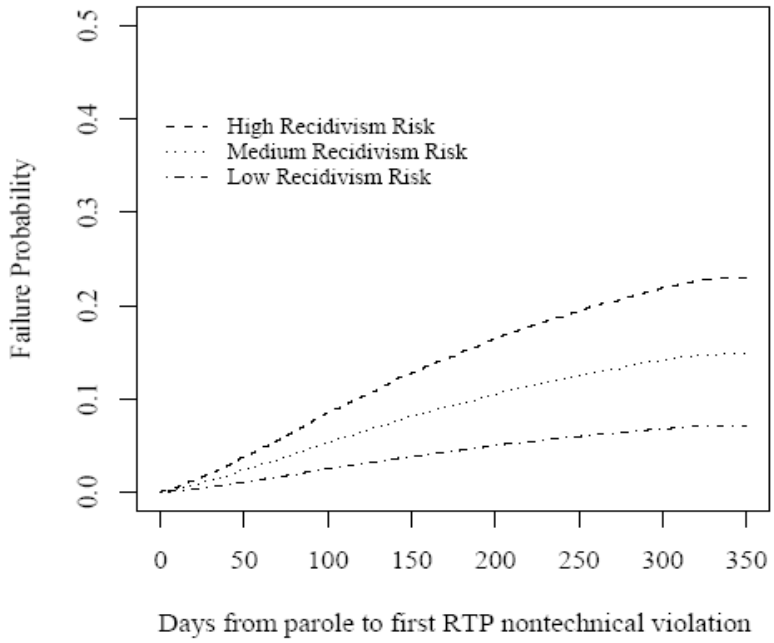
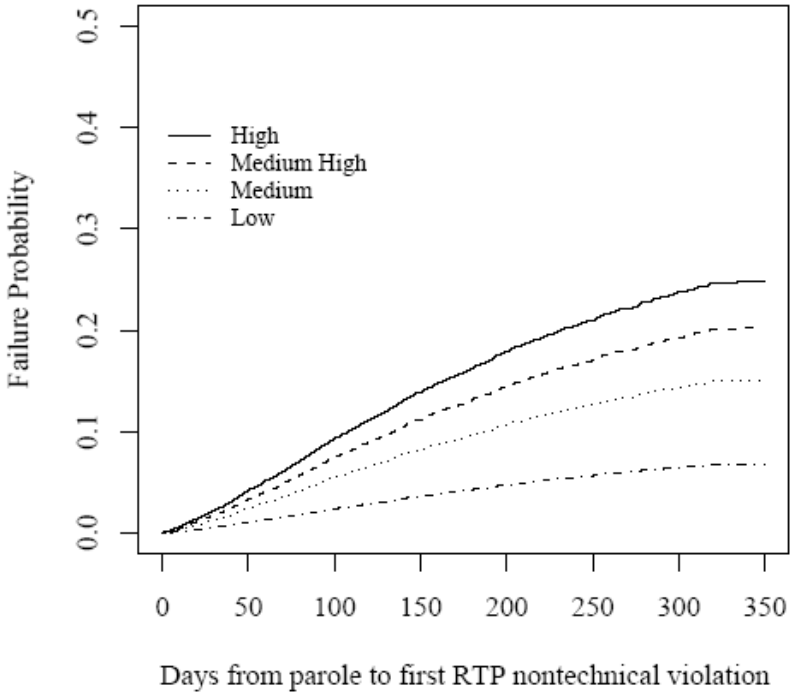


Chart 4.4.2: Predicted Cumulative Incidence of Nontechnical Violation Returns to Prison (RTP) for the Risk Levels of the Matrix-R.



Results for the *Recidivism Risk* scale from cause-specific Cox proportional hazards models are presented in Table 4.4.1. Significant differences between offenders on the different risk levels of the *Recidivism* scale were found regarding the risk of parole failure (here: a return to prison for a nontechnical parole violation). The cause-specific hazard for parole failure is 3.91 for high-risk offenders. This indicates that inmates classified as high-risk have a failure hazard that is about 4 times higher than the hazard for inmates classified as low risk. In comparison, the hazard for the medium category relative to the low category is 2.25.

Table 4.4.1: Proportional Hazard Model Regressing Return to Prison for a Nontechnical Violation on the Recidivism Risk Scale:

Recidivism Risk Level	Coeff.	SE	p-value	Hazard Ratio	Lower 95%CI	Upper 95% CI
Medium	0.813	0.074	< .001	2.25	1.95	2.61
High	1.364	0.068	< .001	3.91	3.42	4.47

Note. The reference category for the test of medium and high risk offenders is the low risk category.

Additional receiver operating characteristics (ROC) analysis for the *Recidivism Risk* scale decile score yields an AUC of .66.

The corresponding model for the *COMPAS Risk Matrix-R* is presented in Table 4.4.2. The model examines the risk of parole failure specified as the cause-specific hazard for parolees to return to prison for a nontechnical violation. The differences between offenders on the risk levels of the *COMPAS Risk Matrix-R* are somewhat larger than the ones found for the *Recidivism Risk* scale, indicating an even stronger risk prediction of the *COMPAS Risk Matrix-R*. Parolees on the high risk level of the *COMPAS Matrix-R* have a 4.7 times higher failure hazard in comparison to parolees on the low risk level. The hazard for the medium category relative to the low category is 2.42.

Table 4.4.2: Proportional Hazard Model Regressing Return to Prison for a Nontechnical Violation on the COMPAS Risk Matrix-R.

Risk Matrix-R	Coeff.	SE	p-value	Hazard Ratio	Lower 95%CI	Upper 95% CI
Medium	0.883	0.080	< .001	2.42	2.07	2.83
Med.High	1.246	0.072	< .001	3.47	3.02	4.00
High	1.542	0.073	< .001	4.67	4.05	5.39

Note. The reference category for the test of medium, medium-high and high risk categories is the low risk category.

Additional receiver operating characteristics (ROC) analysis indicates the area under the curve (AUC) for the *Risk Matrix-R* model presented above is .66.

These findings clearly support the predictive validity of the *Recidivism Risk* scale and the *COMPAS Risk Matrix-R*. While the AUC indicators are moderate, the proportional hazards differ substantially and significantly between the risk levels of both, the *Recidivism Risk* scale

and the *COMPAS Risk Matrix-R.* This indicates a high utility of the scales in discriminating between high and low risk offenders.

5. Construct Validity

This section will synthesize the accumulating body of evidence pertaining to the construct validity of selected COMPAS risk and needs scales. The empirical findings reviewed in the above sections are integrated with the general findings from the recent literature on the meanings and theoretical importance of risk and need factors.

Construct validity includes many different types of validity and according to some researchers even *criterion-based validity* (as discussed above) can be viewed as a form of construct validity (Judd & Kenney, 1981; Judd & McClelland, 1998). Essentially, a scale has high construct validity when it truly measures the construct of interest and consistently behaves in a manner that is theoretically consistent with this construct. Recent research provides evidence that the COMPAS risk scales and/or their components are related to the construct of interest—recidivism. For the sake of brevity, the construct validity of *other* COMPAS scales is not discussed here; please refer to the COMPAS Technical Manual and Psychometrics Report for a detailed discussion of all COMPAS scales.

Criminal Involvement

The *Criminal Involvement* scale is a component of the *Recidivism Risk* scale and the construct validity of the latter is discussed below. The *Criminal Involvement* scale was also shown to be a good predictor of recidivism in Section 4.2 of this report. Offenders that score high on this scale are at a higher risk to recidivate than their counterparts with low scores. The "hazard" of committing a new crime increases by 27.6% for each additional unit of this scale. Since it represents the number of prior instances of arrest, probation, conviction or jail time, this indicates that with every prior event of this type an offender is 27.6% more likely to recidivate. The concurrent validity of this scale could not be validated, because it includes criminal history indicators. The degree of *Criminal Involvement* has consistently emerged as a major risk factor for predicting ongoing criminal behavior. It is the most important of the major risk factors according to various meta-analyses (Lipsey & Derzon 1998, Gendreau, Little, & Goggin, 1996; Andrews & Bonta, 1994).

Vocational/Educational Problems

Another component of the *Recidivism Risk* scale, the *Vocational/Educational Problems* scale also demonstrated high predictive validity as presented in Section 4.2. This COMPAS scale captures the concept of "legitimate economic opportunity" and is an amalgam of educational attainment, vocational skills, job opportunities, employment stability, and good income. The more general concept of "social achievement" is one of the "big five" risk factors for crime and recidivism in the Gendreau, Little, & Goggin (1996) meta-analysis. It is closely related to the theory of social capital (Hagan, 1998; Coleman, 1990). Basically, persons with more social capital have higher "life chances" than others who may have very restricted success opportunities. Offenders differ greatly in access to social capital or other resources. Social Capital is somewhat dynamic—it can be built or destroyed. For example, a record of serious

criminal behavior or high school dropout will clearly diminish life chances and social resources, whereas completing a job skills training course or obtaining a GED may increase these chances.

Drug Problems

This scale summarizes indicators whether the offender has had problems with drugs as derived from the offense history, current charges, or the current arrest, in combination with information regarding a drug treatment history or an expressed need for drug treatment. It is a scale transform specifically developed for modeling recidivism. Because of its high predictive validity, the *Drug Problems* scale became a component of the COMPAS *Recidivism Risk* scale (where prior arrests/convictions for *drug possession* and/or *drug trafficking* were excluded however). It's strong predictive utility has been demonstrated throughout this report with results for the scale itself (Table 4.2.2) as well as results for drug problems as a component of the *Recidivism Risk* scale as discussed below. Numerous published research studies have established that substance abuse is a significant risk factor for both general criminal behavior as well as violent behavior. Substance abuse also emerged as one of the major risk factors in the meta-analysis of Gendreau, Little, & Goggin (1996).

Recidivism Risk Scale and COMPAS Matrix-R

The *Recidivism Risk* scale and *COMPAS Matrix-R* were developed to predict re-offending subsequent to the COMPAS assessment date. They are composed of COMPAS items selected through diagnostic modeling strategies. Whereas scale reliability and coherence were emphasized for developing the other COMPAS scales, predictive validity was clearly emphasized for developing the *Recidivism Risk* scale and the *COMPAS Matrix-R*. Both aim to predict which offenders will commit crimes subsequent to their initial COMPAS screening date. Their predictive validity is central to this report and discussed in the conclusions below.

6. Conclusions and Future Research

Major findings of this report are summarized along the dimensions of reliability and validity that have been examined and evaluated.

Reliability: Findings of good reliability were replicated across several sites and different offender populations with few exceptions. Generally the findings were satisfactory, indicating that the COMPAS scales have high internal consistency and COMPAS is administered in a consistent and reliable manner at all sites included in this comparison.

Concurrent Validity: Correlational patterns commonly found in COMPAS psychometric reports were verified as being rather robust. Overall, the observed relationships between the COMPAS subscales and criminal history indicators data provided evidence of the criterion-concurrent validity of the scales. Their general consistency across different sites and datasets, adds additional empirical support to this conclusion.

Predictive Validity: The COMPAS risk scales are valid tools for determining eligibility for community corrections programs. The offender risk levels clearly predicted program failure.

The predictive utility of several of the COMPAS subscales regarding the risk of recidivism has also been demonstrated in a New York probation sample. Logistic regression confirmed that offenders scoring high on the *Criminal Involvement, Vocational/ Educational, and Drug Problems* scale, are at a substantially higher risk to recidivate. This finding led to the selection of these scales as components of the *Recidivism Risk* scale.

Further evaluations by means of Receiver Operator analysis confirmed the high predictive validity of the *Risk of Recidivism* scale in the New York sample. AUCs for the scales' accuracy in discriminating between recidivists and non-recidivists were similar in magnitude to those obtained by other researchers in the field (Cottle, Lee, & Heilbrun, 2001; Gran, Belfrage, & Tengstrom 2000; Quinsey et al., 1998); but higher than comparable fourth generation risk assessment instruments (Barnoski & Aos 2003). Cross validation in a different New York sample confirmed this assumption with AUCs only slightly smaller than those in the scale construction sample. Therefore the COMPAS *Recidivism Risk* prediction model can be considered robust. Moreover, it can be generalized to other offender populations with little loss of predictive accuracy.

Additional findings in a California sample of parolees examining the utility of the *Recidivism Risk* scale and the *COMPAS Risk Matrix-R* for predicting parole failures clearly support their predictive validity. While the AUC indicators in this sample are rather low, the proportional hazards differ substantially and significantly between offenders on the risk levels of both, the *Recidivism Risk* scale and the *COMPAS Risk Matrix-R,* indicating the scales' effectiveness in discriminating offender probabilities of parole failure. This finding is particularly remarkable because the *Recidivism Risk* scale has been developed in a probation sample and had not been evaluated in a parole population.

Given that instrument validation is an ongoing process, countless further tests and models could be applied to further examine the concurrent and predictive validity of the COMPAS risk scales. The present overview, however, summarized encouraging evidence that the main COMPAS risk scales, *Recidivism Risk* and the *Risk Matrix-R,* perform moderately well in predicting recidivism in a variety of different criminal justice populations. Our next steps to further confirm these findings are additional outcome validation studies in larger samples and with longer follow-up periods. These analyses will be conducted at several different sites including Michigan and New York.

Moreover, a basic or "global" COMPAS will be developed as an attempt to accommodate practitioners' need for a more concise assessment instrument that is nevertheless theoretically and empirically informed. The comprehensiveness of COMPAS, its ability to distinguish different types of offender populations and recidivism risks, and its flexibility in customization will be strengthened as additional options are activated as needed.

References

Andrews, D. A., & Bonta, J. (1994). The psychology of criminal conduct. Cincinnati: Anderson Press.

Aos, Steve & Barnoski, Robert (2003). Washington's Offender Accountability Act: An Analysis of the Department of Correction's Risk Assessment. Washington Institute for Public Policy: Document Number: 03-12-1202

Austin, James (2006) How much Risk Can we Take? Federal Probation, Vol. 70, No. 2, 58-63

Brennan, Tim, Dieterich, William (2007). A Test of the Validity of the COMPAS Risk Scales: Predicting Return to Prison in a Large Parole Sample. Unpublished Report. Northpointe Institute for Public Management. Traverse City, Michigan.

Coleman, J. S. (1990). Foundations of social theory. Cambridge, Mass: Harvard University Press.

Cottle, Cindy C., Lee, Ria J., Heilbrun, Kirk (2001). The Prediction of Criminal Recidivism in Juveniles. Criminal Justice and Behavior, Vol.28 No.3, 367-394.

Farrington, D. P., Jolliffe, D., Loeber, R., Stouthamer-Loeber, M., & Kalb, L. M. (2001). The concentration of offenders in families, and family criminality in the prediction of boys' delinquency. Journal of Adolescence, 24, 579–596.

Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association, 94 (446), 496–509.

Gendreau, P., Goggin, C., & Little, T. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! Criminology, 34, 575-607.

Grann, M., Belfrage, H., & Tengstrom, A. (2000). Actuarial assessment of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20. Criminal Justice and Behavior, 27, 97-114.

Hagan, J. (1998). Life course capitalization and adolescent development. In R. Jessor (Ed.), New perspectives on adolescent risk behavior, 499–517. New York: Cambridge University Press.

Hosmer Jr., D. W., & Lemeshow, S. (1998). Applied survival analysis: Regression modeling of time to event data. New York City: John Wiley.

Jones, P.R., Risk Prediction in Criminal Justice, in A. T. Harland, ed., "Choosing Correctional Options That Work," 33-68. Thousand Oaks, CA: Sage, 1996.

Judd, C. M., & Kenney, D. A. (1981). Estimating the effects of social interventions. New York: Cambridge University Press.

Judd, C. M., & McClelland, G. H. (1998). Measurement. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), The handbook of social psychology, 180–232. Boston, MA: McGraw-Hill.

Lipsey, M.W., & Derzon, J.H. (1998) Predictors of violent and serious delinquency in adolscence and early adulthood: A synthesis of longitudinal research. In: Loeber, R. & Farrington, D.P. (Eds.), Serious and violent juvenile offenders: Risk factors and successful interventions, 86-105. Thousan Oaks, CA: Sage.

Moffitt, T. E. (1993). Adolescence-limited and life-course persistent antisocial behavior: A developmental taxonomy. Psychological Review, 100 (4), 674–701.

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

Stouthamer-Loeber, M., Loeber, R., Wei, E., Farrington, D. P., & Wikstrom, P. H. (2002). Risk and promotive effects in the explanation of persistent serious delinquency in boys. Journal of Consulting and Clinical Psychology, 70 (1), 111–123.

Tuma, N.B., Hannan, M.T. (1984), Social Dynamics. Models and methods. New York: Academic Press.

Venables, W. N., & Ripley, B. D. (1999). Modern applied statistics with S-Plus. New York City: Springer-Verlag.

APPENDIX A—Scale Descriptions

Table A1: Criminal Involvement Scale Items (criminv).

Items	Short Description (Response Categories)
n.jails	How many times has the offender been jailed?
	(0=0, 1=1, 2=2, 3=3-7, 4=8-12, 5=13+)
n.prev.convict	How many times has the offender been convicted before?
	(0=0, 1=1, 2=2, 3=3-4, 4=5-10, 5=11+)
n.prev.arrest	How many times has the offender been arrested before?
	(0=0, 1=1, 2=2-3, 3=4-5, 4=6+)
n.probations	How many times has the offender been on probation?
	(0=0, 1=1, 2=2, 3=3, 4=4-5, 5=6+)

Table A2: Vocational/Educational Problems, Scale Items (voced).

Items	Short Description (Response Categories)
high.school	Did offender complete 12th grade?
	(1=yes, 2=no)
expelled	Was offender ever suspended or expelled from school?
	(2=yes, 1=no)
grades.hs	What were the offender's usual grades in high school?
	(A=1, B=2, C=3, D=4, F=5)
job	Does the offnder currently have a job?
	(1=yes, 2=no)
skill	Have skill, trade or profession to usually find work?
	(1=yes, 2=no)
job.lastyear	How much work or school the last 12 months?
	(1=12 months full time, 2=12 months part time,
	3=6+ months full time, 4=less than 6 mos. PT/FT)
fail.or.repeat.grd	Did the offnder fail or repeat a grade level?
	(2=yes, 1=no)
need.training	Feel need more training in a new job or career skill?
	(2=yes, 1=no)
wages.above.min	How hard to find a job above minimum wage?
	(1=easier, 2=same, 3=harder, 4=much harder)
chance.success.work	How would they rate their chance of being successful?
	(1=good, 2=fair, 3=poor)
haveempschool	Does offender have verified local employer or school?
	(1=yes, 2=no)

Table A3: Drug Problem Component Items (drgprob5).

Items	Short Description (Response Categories)
d.current	using drugs when arrested for current offense 2=Yes, 1=No
curr.drugp	Current charge drug possession (0=not checked, 1= checked)
curr.drugt	Current charge drug Trafficking (0=not checked, 1= checked)
want.rx.d	Would benefit from treatment for drugs 2=Yes, 1=No
ever.rx.d	Has offender ever been in treatment for drugs 2=Yes, 1=No
drugposs*	Number of previous arrests/convictions for drug possession (0=0, 1=1, 2=2, 3=3+)
drugtraf*	Number of previous arrests/convictions for drug sales (0=0, 1=1, 2=2, 3=3+)

*Note: These items were included in the original scale construction model (New York Probation) but are not a component of the Recidivism Risk scale anymore.

Table A4: History of Violence Scale Items (histviol).

Items	Short Description (Response Categories)
n.prev.vfel	Total prior juvenile/adult felony assault arrests (0, 1, 2, 3, 4, 5+)
assault	Total prior assault (not murder) arrests/convictions (0, 1, 2, 3+)
sex.force	Total prior sex offense with force arrests/convictions (0, 1, 2, 3+)
homicide	Total prior homicide/manslaughter arrests/convictions (0, 1, 2, 3+)
robbery	Total prior robbery arrests/convictions (0, 1, 2, 3+)
weapons.offense	Total prior weapons arrests/convictions (0, 1, 2, 3+)
family.violenc	Total prior family violence arrests/convictions (0, 1, 2, 3+)
frq.injury	Prior times victim had physical injuries (0, 1, 2, 3, 4, 5+)
fights.inmate	Ever writeups for fighting/threatening inmates? (2=yes, 1=no)

Table A5: History of Noncompliance Scale Items (histnonc).

Items	Short Description (Response Categories)
n.prob.rev	How many times probation suspended or revoked? (0, 1, 2, 3, 4, 5+)
n.arrest.on.bail	How many times arrested/charged for crime on pretrial release? (0, 1, 2, 3+)
n.fta	How many times failed to appear on time for court? (0, 1, 2, 3, 4, 5+)
n.rec.prob	How many times arrested/charged while on probation/parole? (0, 1, 2, 3+)