Revalidation of the Women's Risk Needs Assessment:

Pre-Release Results[1]

Final Report

August 2012

Patricia Van Voorhis, Ph.D.
University of Cincinnati

Ashley Bauman, M.S., M.B.A.
University of Cincinnati

Rachel Brushett, M.S.
University of Cincinnati

[1] This research was funded by the National Institute of Corrections under cooperative agreement 09M12GKB3. Points of view or opinions stated in this article are those of the author and do not necessarily represent the official position or policies of the United States Department of Justice or the National Institute of Corrections.

TABLE OF CONTENTS

Executive Summary…… iv

Acknowledgments……. 2

Introduction…… 4

Description of Participating Jurisdictions…..…… 7

Missouri…… 7

Kentucky…….. 8

Ohio…….. 8

Rhode Island…… 10

Methodology……..... 10

Sample Description ……..... 11

Analytic Process……..…….... 17

Offense-Related Outcome Measures……. 18

Results ……..……..... 22

Revalidation of Individual Risk/Need Scales Created During Construction Validation Research……... 22

Revalidation of the Original Cumulative Risk Scales Created During Construction Validation Research……. 35

Revision of the WRNA Scales……. 38

Constructing a Final Stand-Alone Assessment and Trailer……. 54

Changes to the WRNA Assessment…….. 60

Conclusion……..... 63

References …… 66

APPENDICES ……. 68

LIST OF TABLES

Table 1. Frequency and Percent Distribution of Demographic and Offense-Related Background Characteristics of Samples Participating in the Validation of the Stand-Alone WRNA…… 12

Table 2. Frequency and Percent Distribution of Demographic and Offense-Related Background Characteristics of Samples Participating in the Validation of the Supplemental WRNA (WRNA-T)..…… 15

Table 3. Frequency and Percent Distribution of Demographic and Offense-Related Background Characteristics of the Construction and Revalidation Samples…….. 19
Table 4. Follow-up Measures by Time Frame and Site……. 21
Table 5. Bivariate Relationship between Original WRNA Assessment Scales and 6-Month Offense-Related Outcomes, by Site..…….. 24
Table 6. Bivariate Relationship between Original WRNA Assessment Scales and 12-Month Offense-Related Outcomes by Site……. 27
Table 7. Bivariate Relationship between LSI-R and WRNA Assessment Scales and 6-Month Offense-Related Outcomes, by Site……..... 31
Table 8. Bivariate Relationship between LSI-R and WRNA Assessment Scales and 12-Month Offense-Related Outcomes, by Site…… 33
Table 9. Bivariate Relationship between Original WRNA Assessment Scales and 6 and 12 Month Offense-Related Outcomes, by Site…...…… 37
Table 10. Bivariate Relationship between Original WRNA-T Assessment Scales and 6 and 12 Month Offense-Related Outcomes, by LSI-R Sites…..……... 39
Table 11. Bivariate Relationship between Revised WRNA Assessment Scales and 6-Month Offense-Related Outcomes, by Construction and Revalidation Samples... 41
Table 12. Bivariate Relationship between Revised WRNA Assessment Scales and 12-Month Offense-Related Failures, by Construction and Revalidation Samples….. 44
Table 13. Bivariate Relationship between Revised WRNA Assessment Scales and 6 and 12-Month Recidivism, across Construction and Revalidation Samples…..…… 57
Table 14. Bivariate Relationship between Revised WRNA Assessment Scales and 6 and 12-Month Recidivism, across Jurisdictions……..……. 58
Table 15. Bivariate Relationship between LSI-R and Revised WRNA-T and 6-Month Recidivism, across Jurisdictions……. 62

LIST OF APPENDICES

Appendix A: Original Women's Risk Needs Assessment 68
Appendix B: Original WRNA Scoring Form 94
Appendix C: State Specific Findings for the Revised WRNA 99
Appendix D: Alpha Measures of Internal Consistency 110
Appendix E: Revised Women's Risk Needs Assessment 112
Appendix F: Revised WRNA Scoring Form 135
Appendix G: Revised Women's Risk Needs Assessment – Trailer (WRNA-T) 140
Appendix H: Revised WRNA-T Scoring Form 158

Revalidation of the Women's Risk Needs Assessment:
Pre-Release Results[2]
Final Report

August 2012

Executive Summary

Patricia Van Voorhis, Ph.D.
University of Cincinnati

Ashley Bauman, M.S., M.B.A.
University of Cincinnati

Rachel Brushett, M.S.
University of Cincinnati

In order to facilitate appropriate treatment planning and risk management for women offenders, the National Institute of Corrections (NIC) and the University of Cincinnati (UC) engaged in a series of cooperative agreements that resulted in the development of two types of gender-responsive assessments. The work began in 1999 with a pilot study in the Colorado Department of Corrections and continued with three projects in Maui, Minnesota, and Missouri. The first assessment, called the Women's Risk Needs Assessment – Trailer (WRNA-T) (or "the trailer") was designed to supplement existing dynamic risk/needs assessments such as the Level of Service Inventory-Revised (LSI-R) (Andrews & Bonta, 1995) and the Northpointe COMPAS (Brennan, Dieterich, & Oliver, 2006) The second, the Women's Risk Needs Assessment (WRNA), was an assessment that could be used on its own, as a "stand-alone," dynamic, risk/needs assessment, comprised of both gender-neutral and gender-responsive scales.

[2] This research was funded by the National Institute of Corrections under cooperative agreement 09M12GKB3. Points of view or opinions stated in this article are those of the author and do not necessarily represent the official position or policies of the United States Department of Justice or the National Institute of Corrections.

Extensive literature searches and focus groups with correctional administrators, treatment practitioners, line staff, and women offenders informed both of the assessments. Both instruments contained an interview and a self-report survey.

The full instrument, and most of the questions now contained on the trailer, was developed by members of the Women's Issues Committee of the Missouri Department of Corrections (MDOC) in collaboration with researchers at the University of Cincinnati and key staff from the National Institute of Corrections. This construction process also benefitted from the expertise of substance abuse specialists, psychologists and other mental health professionals on staff with MDOC. The construction validation studies also produced different versions of the WRNA and the WRNA-T for specific types of correctional populations, because it was discovered that the predictive validity of both the gender neutral and the gender-responsive variables varied by correctional settings, e.g., prerelease, probation, and prisons.

This report presents findings from a second cooperative agreement between the University of Cincinnati and the National Institute of Corrections. The present study was begun in 2009. Since the earlier assessments were created through construction validation, a key goal of the present study was to revalidate the original versions on new samples of offenders to determine the level of shrinkage in predictive validity from the construction to revalidation studies. Additionally, the 2009 cooperative agreement sought to refine several of the dynamic risk/needs scales in order to further improve predictive validity. In doing so, this research tested a number of new items that allowed for the exploration of their potential contributions to a revised assessment. Of course, creating a revised assessment also required another revalidation. To accomplish that, the new studies furnished larger samples than produced by the construction validation research and afforded an opportunity to partition the combined samples into construction and revalidation samples. The present report focuses on the Pre-Release WRNA and WRNA-T, the assessments intended for use in reentry planning for women released from prison. Two additional reports will provide similar findings for female inmates and probationers.

The rationale for seeking to improve the Pre-Release WRNA reflected the fact that the original tool was developed on a small Missouri sample (N=149). In contrast, the present study secured a sample of 403 participants across three sites, Missouri, Ohio, and Kentucky, to test gender-neutral variables and 626 participants across 4 sites, Missouri, Ohio, Kentucky, and Rhode Island, to test gender-responsive variables. The following analytical steps were employed:

1. Individual risk/need scales developed during the original study (N=149) were tested, through analysis of correlations (Pearson's r and AUC values) with outcome measures. These tests involved the same items and scoring protocols resulting from the 2004-2008 construction validation study. Analyses were run twice, once for the full assessment (WRNA sites) and once for the trailer assessment (WRNA-T sites).

2. The original total risk/needs score (developed through construction validation research), including risk levels, was tested on the research samples for the present study (Missouri, Ohio, and Kentucky). Additionally, the cumulative WRNA-T scales were added to the LSI-R for the Ohio and Rhode Island samples and tested for predictive validity.

3. The current study collected data on a number of new test items to determine whether they improved the predictive validity of individual domain/need scales. Items were tested on a split-half sample of all prerelease sites studied. The total sample was divided into a construction sample (N=201 for gender-neutral scales and 314 for gender-responsive scales) and a revalidation sample (N=202 for gender-neutral scales and 312 for gender-responsive scales). The construction and revalidation samples were drawn through a systematic random selection process where every other case from the total pool of participants was selected for the construction sample, and the remaining cases were reserved for the revalidation sample. Items were developed on the construction validation sample and retested (confirmed) on the revalidation sample.

4. Because both the construction and revalidation samples were small, another analysis was conducted which tested the new scales for each state sample. Scales found to be predictive in the construction validation study but not in the revalidation study, could nevertheless be retained for the final assessment if they were found to be predictive in two or more of the state samples. This rather unusual procedure accommodates some degree of sample specificity which occurred for both the WRNA and the LSI-R risk/need scales. That is, there was a tendency for a given risk/need domain to be predictive in some samples and not in others, regardless of the assessment used (LSI-R or WRNA).

5. Selection of a final risk/needs stand-alone and trailer scales considered both the results for the construction and revalidation samples (step 3) and the state- specific findings (step 4). The WRNA and the WRNA-T were developed in the construction validation sample, retested on the revalidation sample, and then tested for specific sites.

Offense-Related Outcome Measures

Most participants were followed up for 12 months, and results were reported at a 6 month interval and for the entire 12 months. Because parolees could fail (recidivate) in a variety of ways, a number of outcome measures were examined: a) NEW ARRESTS (Y/N); 2) NEW CONVICTIONS (Y/N); 3) RETURNS TO PRISON (through technical or law violations or new arrests/convictions); 4) ANY OFFENSE-RELATED FAILURE (e.g., new arrests/convictions as well as behavior which could have been processed as a violation but through officer discretion or agency policy was nevertheless recorded as a violation), and 5) ANY FAILURE (any of the above). As can be seen from Table 1, there was considerable variation across sites as to how offenders failed. In Missouri, for example, there were very

Table 1. Follow-up Measures by Time Frame and Site

Site	Arrests		Conv.		Ret Pris[a]		Offense Fail		Any Fail	
	N	%	N	%	N	%	N	%	N	%
6 Month Follow-up										
Missouri (N=195)	14	7.2	0	0.0	17	8.7	30	15.4	99	50.8
Kentucky (N=36)	8	22.2	4	11.1	8	22.2	9	25.0	9	25.0
Ohio (N=172)	40	23.3	21	12.2	12	7.0	51	29.7	62	36.0
Ohio-WRNA-T (N=137)	30	21.9	16	11.7	7	5.1	36	26.3	42	30.7
Rhode Island-WRNA-T (N=222)	44	19.7	24	10.8			50	22.4	98	43.9
12 Month Follow-up										
Missouri (N=187)	24	12.8	6	3.2	37	19.8	61	32.6	125	66.8
Kentucky (N=35)	14	40.0	7	20.0	14	40.0	15	42.9	16	45.7
Ohio (N=169)	57	33.7	32	18.9	23	13.9	71	42.0	85	50.3
Ohio-WRNA-T (N=134)	45	33.6	25	18.7	15	11.2	53	39.6	62	46.3
Rhode Island-WRNA-T (N=210)	69	32.9	39	18.6			76	36.2	121	57.6

[a] Data on returns to prison were not available for Rhode Island.

few arrests and even fewer convictions[3], and returns to prison were through a violation process. As such measures of arrests and convictions are not as meaningful for that sample as they were for Ohio, Kentucky, or Rhode Island. The various Kentucky follow-up measures appeared to be redundant. It is likely that the summary measure of ANY OFFENSE-RELATED FAILURE is the most comparable across sites.

Base rates at 6 months and for some site-specific follow-up measures were problematic. Most notably, in contrast to the 2004 to 2008 construction validation research, the study was not funded for a 24 month follow-up time frame, which likely would have improved base rates.

Results

Revalidation of Individual Risk/Need Domains Created During Construction Validation Research

Results for the tests of the relationship between risk/need domains and offense-related outcomes varied by sample and outcome measure. At 6 months, risk/need domains that were predictive (on at least one of the outcome measures) for all three sites testing the WRNA stand alone assessment were antisocial attitudes, housing safety, antisocial friends, child abuse, sexual abuse, current/recent substance abuse, and parental difficulties. Other measures, such as criminal history, employment financial, anger, adult victimizations, physical abuse, parental stress, and self-efficacy were associated with outcomes in some samples but not others. By 12 months, only antisocial associates and current/recent substance abuse predicted in all settings. Results were also favorable for criminal history (Ohio), employment/financial (Missouri and Ohio), adult victimization (Kentucky and Ohio), substance abuse history (Kentucky), depression (Ohio), relationship satisfaction (Kentucky), relationship dysfunction (Kentucky and Missouri), family support (Kentucky and Missouri), and parental stress (Missouri and Ohio).

In two sites which tested the LSI-R and the WRNA-T scales (Ohio and Rhode Island), a number of the LSI-R scales were predictive on a sample-specific basis as well. These included antisocial attitudes, accommodations, and leisure/recreation at 6 months, and financial, leisure recreation, and substance abuse history at 12 months. Many of the gender-responsive variables, including employment/financial, housing safety, anger, psychosis, child abuse, parental involvement, and self-efficacy, at 6 months, and employment/financial, anger, depression, adult abuse, sex abuse, parental involvement, at 12 months, were predictive for both samples. Others scales tended to be predictive in one

[3] In fact, no convictions were noted in the Missouri follow-up data during the first 6 months.

sample but not the other regardless of follow-up time frame. Family support, family conflict, and the relationship scales were not predictive.

The results likely reflect a number of differences across jurisdictions, including base rates on follow-up measures, interviewer skill sets, programming resources, and cultural, ethnic and other offender background characteristics. The fact that the results were especially strong for the Kentucky sample, especially for the gender responsive variables, reflects one of the challenges in administering dynamic risk/needs assessments to pre-release samples. Results for dynamic assessments administered in prisons are likely to change upon release to communities. Kentucky was the only pre-release settings where assessments were administered after release, after the offenders had returned to their home communities.

Another pattern in these findings, suggests that findings were better in those sites where the assessment was actually being used for treatment planning and case management. For example, Rhode Island and Kentucky evidenced especially strong results. The assessments administered in Ohio and Missouri were for research purposes. Missouri was experiencing problems with staff buy-in and Ohio assessments were administered by research personnel at the University of Cincinnati.

Revalidation of the Original Cumulative Risk Scales Created During Construction Validation Research (2004-2008).

Table 2 shows the results of the revalidation of the original stand-alone risk scale in samples where the full WRNA was tested. These cumulative total scales used the same cut-points, need domains, and scoring conventions as used in the original study. The Kentucky site achieved correlations and AUC values that were in a favorable range. At 6 months correlations ranged from .25 to .35 across outcome measures, and from .27 to .38 on 12 month outcomes. Findings for Missouri and Ohio were much less impressive and in need of improvement, but they were typical of results seen at 12 months in the earlier research. Table 2 also shows findings from a second analysis which was conducted on only the gender responsive cumulative scales (WRNA-T), scales that would be used in a trailer if a jurisdiction chose not to use the stand-alone assessment. These must be interpreted with caution, because this cumulative gender-responsive scale is not intended to be used as a full instrument as some have suggested (Andrews & Bonta, 2010). The point here is to show that as a block of factors, these scales are significantly related to outcomes. However, the scale is not intended to reach the strengths of association seen with full scales (gender-neutral plus gender-responsive). Table 2 notes in fine print several 6 and 12 month findings for Kentucky. These were equal in magnitude to findings for other sites, but did not reach significance due to the small sample size and limited statistical power.

Table 2. Bivariate Relationship between Original WRNA Assessment Scales and 6 and 12 Month Offense-Related Outcomes, by Site.

	Arrests		Conviction		Return to Prison		Offense-Related Failure		Any Failure	
Scale	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**
6 Month Follow-up for Original Stand-Alone WRNA										
Missouri Original WRNA (N=195)					.15**	.64			.20***	.62
Missouri-Levels					.13**	.60	.12**	.57	.25***	.62
Kentucky Original WRNA (N=36)	.35**	.78			.25*	.71	.32***	.75	.32***	.75
Kentucky-Levels	.33**	.70	.22*	.67			.26*	.65	.26*	.65
Ohio Original WRNA (N=172)	.21***	.64	.19***	.63	-.10*		.16**	.60	.14**	.58
Ohio-Levels	.18***	.61					.15**	.58	.12*	.56
6 Month Follow-up Original WRNA-T Scales										
Missouri WRNA WRNA-T (N=195)					.12**	.61	.10*	.55	.21***	.62
Kentucky Island WRNA -T (N=36)	.24**	.68	.16	.65	.12	.59	.19	.64	.19	.64
Ohio WRNA-T (N=172)	.25***	.66	.12*	.60			.20***	.62	.15**	.58
12 Month Follow-up for Original Stand-Alone WRNA										
Missouri Original WRNA (N=187)	-.10*		.18***	.76	.19***	.64	.15**	.59	.17***	.60
Missouri-Levels			.18***	.72	.18***	.61	.14**	.57	.14***	.57
Kentucky Original WRNA (N=35)	.38***	.73	.27*	.71	.29**	.68	.36**	.73	.31**	.69
Kentucky-Levels	.35***	.68	.29**	.88	.24*	.62	.30**	.65	.25*	.62
Ohio Original WRNA (N=169)	.23***	.63	.20***	.63			.20***	.61	.15**	.58
Ohio-Levels	.21***	.61	.17***	.61			.21***	.69	.13***	.56
12 Month Follow-up for Original WRNA-T Scales										
Missouri Original WRNA-T (N=187)	-.16***				.14**	.61			.18**	.61
Kentucky Original WRNA-T (N=35)	.13	.58	.28*	.70						
Ohio Original WRNA -T (N=169)	.21***	.62	.13**	.60			.19***	.61	.15**	.59

***p≤.01
**p≤.05
*p≤.10

Results for settings testing the WRNA-T (trailer) and adding it to LSI-R cumulative scales are shown in Table 3, below. Again, these use the scales developed during the construction validation study conducted between 2004 and 2008. The contribution of the trailer items (partial correlation) was statistically significant at 6 months in 8 of 10 tests conducted.

By 12 months, however, the predictive validity of the LSI-R improved for the Ohio sample. While the relationship between the WRNA-T and the two outcomes measures was significant in Ohio and made a very modest improvement to the LSI-R in 3 of 5 tests, the partial correlation was not significant. This was not the case for Rhode Island where across all outcome measures, the trailer made an important contribution to the predictive validity of the LSI-R. The difference between the results for Ohio and Rhode Island may also reflect the truncated distribution of the Ohio sample which included very few low risk offenders.

Revision of the WRNA Scales

Improvement of the assessment scales was achieved through the development and testing of new items on a construction validation sample and then revalidating those items. Presumably, the revalidation of the new scales will reduce concerns for the need of another revalidation study. The new items and item analyses were conducted on 19 of 28 risk/need domains. These analyses resulted in improvements to the predictive validity of 14 of the 19 domains.

The final stand-alone WRNA scale consisted of the following individual risk/need domains:

Criminal history
Employment/financial
Housing safety
Antisocial friends
Anger
Depression (collapsed)
Psychosis
Child abuse (interview scales)
Adult abuse (interview scales)
Sexual abuse (interview scales)
Current substance abuse

Table 3. Bivariate Relationship between Original WRNA-T Assessment Scales and 6 and 12 Month Offense-Related Outcomes, by LSI-R Sites.

	Arrests		Conviction		Return to Prison		Offense-Related Failure		Any Failure	
Scale	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**
6 Month Follow-up										
Ohio LSI-R (N=137)							.13*	.58	.11*	.57
Ohio WRNA-T (N=137)	.21***	.64					.20***	.63	.17**	.61
Ohio LSI-R + WRNA-T (N=137)	.17**	.60					.18**	.61	.16**	.59
Partial corr	.19**						.17**		.15**	
Rhode Island LSI-R (N=223)	.11*	.57	.10*	.58			.10*	.56	.16***	.58
Rhode Island WRNA-T (N=223)	.16***	.61	.18***	.66			.22***	.64	.17***	.59
Rhode Island LSI-R + WRNA-T (N=223)	.13**	.57	.13**	.60			.15**	.58	.18***	.59
Partial corr.	.13**		.15***				.20***		.09*	
12 Month Follow-up										
Ohio LSI-R (N=134)	.16**	.59			.14**	.65	.23***	.64	.19***	.61
Ohio LSI-R WRNA-T (N=134)	.13*	.58					.17**	.60	.14**	.58
Ohio LSI-R + WRNA-T (N=134)	.18**	.60			.12*	.64	.25***	.65	.20***	.62
Partial corr.										
Rhode Island LSI-R (N=210)	.14**	.57	.14**	.59			.18***	.60	.21***	.62
Rhode Island WRNA-T (N=210)	.26***	.65	.25***	.68			.29***	.67	.21***	.62
Rhode Island LSI-R + WRNA-T (N=210)	.19***	.59	.19***	.62			.23***	.62	.23***	.63
Partial corr.	.22***		.21***				.23***		.10*	

***p≤.01
**p≤.05
*p≤.10

Relationship dysfunction
Parental stress (collapsed)

Additionally, the following strengths were subtracted from the total scale:

Family support (collapsed)
Self-efficacy (collapsed)

Table 4 shows results of tests conducted on the construction and revalidation samples. Correlations for the construction validation study, as expected, were high, especially for the collapsed levels, where correlations met or surpassed a standard of r=.27 for 7 of the ten tests. The relationship between the stand alone scale and ANY OFFENSE-RELATED FAILURE, the measure considered to be the most uniform measure of outcome across sites, (Pearson's r) was *.29, p ≤. 01* at 6 months, and r=.32*, p ≤. 01* at 12 months. On revalidation, some shrinkage in predictive validity was seen, however, the 12 month results for ANY OFFENSE-RELATED FAILURE remained strong.

AUC values seldom reached a standard of .70 in either the construction or revalidation tests, except on measures of CONVICTIONS at 12 months. This may be attributable to the truncated follow-up period. Results for the analysis of 12 month outcomes are stronger than those for 6 months. This phenomenon is likely to be the result of improvements in outcome base rates over time. In most recidivism research such improvements continue until the 2 or 3 year follow-up point. As such, two years of follow-up appears to be the scientific standard for community recidivism studies.

Selection of scales for the WRNA-T followed a similar procedure, except scales deemed to be redundant to similar scales on the LSI-R (e.g., family support) were not included unless they provided a gender-responsive definition of a similar scale, e.g., housing safety and employment/financial. The cumulative scale for the WRNA-T consisted of the following risk/need domains:

Employment/financial
Housing safety
Anger
Depression (collapsed)
Psychosis
Child abuse (interview scales)
Adult abuse (interview scales)
Sexual abuse (interview scales)
Relationship dysfunction

Table 4. Bivariate Relationship between Revised WRNA Assessment and 12-Month Recidivism, across Construction and Revalidation Samples.

	Arrests		Conv.[a]		Ret Pris[b]		Offense Fail		Any Fail	
Scale	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**
Construction Validation Sample 6 Months										
WRNA Stand-alone Assessment (N=201)	.22***	.66	.33***	.79			.23***	.66	.23***	.64
Levels (N=201)	.28***	.69	.29***	.74	.09*	.60	.29***	.69	.27***	.64
WRNA-T: Trailer (N=312)	.17***	.61	.34***	.74			.21***	.62	.19***	.60
Construction Validation Sample: 12 Months										
WRNA Stand-alone Assessment (N=194)	.24***	.65	.25***	.70	.15**	.64	.29***	.67	.24***	.64
Levels (N=194)	.29***	.67	.24***	.68	.16***	.62	.32***	.67	.27***	.64
WRNA-T: Trailer (N=302)	.19***	.60	.26***	.70			.20***	.61	.18***	.59
Revalidation Sample: 6 months										
WRNA Stand-alone Assessment (N=201)	.18***	.62	.22**	.67	.12**	.59	.22***	.63	.21***	.61
Levels (N=201)	.16***	.60	.15*	.60	.13*	.60	.23***	.62	.24***	.63
WRNA-T: Trailer (N=312)	.20***	.62	.13**	.60			.19***	.61	.16***	.58
Revalidation Sample: 12 months										
WRNA Stand-alone Assessment (N=201)	.20***	.62	.29***	.74	.16***	.61	.26***	.65	.18***	.61
Levels (N=201)	.18***	.61	.26***	.70	.18***	.62	.28***	.65	.22***	.62
WRNA-T: Trailer (N=298)	.18***	.60	.23***	.67			.21***	.61	.14***	.68

[a] Missouri cases are omitted from analysis of 6 month conviction data. No convictions had occurred by that point in time.
[b] Rhode Island cases are not included in any analyses of returns to prison. Data on returns to prison were not available.
***p≤.01
**p≤.05
*p≤.10

Parental stress (collapsed)

Self-efficacy (collapsed) was subtracted from the total.

As shown in Table 5, results for the WRNA-T, alone (second row for each site), were somewhat better for Rhode Island than Ohio, but in both cases, addition of the WRNA-T items to the LSI-R improved the predictive validity of the LSI-R. Moreover, partial correlations show that for all tests conducted, the improvement attributable to the WRNA-T was statistically significant. It should be noted that the Ohio data provided very few low risk offenders, thereby truncating the distribution of the assessment data.

Table 5 also shows a number of instances where correlations for the trailer plus the LSI-R (third row) are lower than those for the trailer alone. This was not intended, but likely occurred because associations between the LSI-R and the outcome variables were not sufficiently strong. Although this was unanticipated, it implicates the quality of the LSI-R at the sites where the LSI-R data were collected. The WRNA interviewers were trained immediately prior to the administration of the WRNA assessment while at least some of the LSI-R interviewers may have been due to receive retraining. In addition, the Ohio LSI-R data was extracted from institutional record data which may have been somewhat dated.

It can also be seen that correlations between collapsed levels and outcome measures were not as high as those for the total LSI and WRNA-T. This reflected some difficulty in setting uniform level cut-points for the two sites. That is, they could have been optimized for each site, but they were not. As a result, the score sheet for the revised WRNA-T recommends that sites re-examine the cut-points and optimize them for their own jurisdictions.

Changes to the WRNA Assessment

The present study has resulted in a streamlined assessment that reduces the number of interview questions from 145 to 116. Survey questions have been reduced from 69 to 36. Most importantly, the current study shows that the revised WRNA and the WRNA-T work across settings. Earlier versions were "fit" to sites, because it was not possible to develop an assessment on a large sample of participants at the outset. Data were not available at a single point in time, so that a large construction and revalidation sample could not be conducted first and then applied to specific sites as in the present study. This study has greatly improved the uniformity of the tool. Sites will vary somewhat in terms of the specific risk/need domains that are predictive, and that is typical of dynamic risk/need assessments, but the total scale is predictive in all sites and cut-points for larger risk-need domain scales, including the total risk scale, are uniform.

Table 5. Bivariate Relationship between LSI-R and Revised WRNA-T and 6-Month Recidivism, across Jurisdictions.

	Arrests		Conviction		Return to Prison		Offense-Related Failure		Any Failure	
Scale	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**
6 Month Follow-up for LSI-R and WRNA-T										
Ohio LSI-R (N=137)							.13*	.58	.11*	.57
Ohio WRNA T N=137)	.21***	.61	.23***	.63			.19**	.59	.15**	.57
Ohio LSI-R + WRNA-T (N=137)	.18**	.61	.16**	.61			.19**	.61	.16**	.59
Partial Correlation	.18**		.23***				.15**		.13*	
LSI-WRNA-T Levels	.12*	.56					.14*	.57	.13*	.57
Rhode Island LSI-R (N=222)	.11*	.57	.10*	.58			.10*	.56	.15***	.59
Rhode Island WRNA-T N=222)	.19***	.58	.20***	.66			.24***	.60	.19***	.58
Rhode Island LSI-R +WRNA T (N=222)	.16**	.57	.16***	.63			.18***	.59	.20***	.59
Partial Correlation	.16***		.18***				.22***		.12*	
LSI-WRNA-T Levels	.16***	.59	.13***	.60			.18***	.60	.18***	.59
12 Month Follow-up for the LSI-r and WRNA-T										
Ohio LSI-R (N=134)	.16**	.60		.54	.14**	.66	.23***	.65	.19**	.62
Ohio WRNA-T N=134)	.17**	.58	.20***	.61			.19**	.59	.18**	.58
Ohio LSI-R + WRNA-T (N=134)	.20***	.61	.14**	.59			.26***	.64	.22***	.62
Partial Correlation	.12*		.20**				.12*		.12*	
LSI-WRNA-T Levels	.16**	.58			.12*	.60	.22***	.61	.20***	.61
Rhode Island LSI-R (N=209)	.14**	.57	.14**	.59			.18***	.60	.21***	.62
Rhode Island WRNA- T N=209)	.22***	.60	.28***	.68			.27***	.63	.20***	.59
Rhode Island LSI-R +WRNA-T (N=209)	.20***	.59	.23***	.65			.25***	.62	.24***	.62
Partial Correlation	.18***		.25***				.21***		.11**	
LSI-WRNA-T Levels	.20***	.60	.19***	.62			.23***	.62	.22***	.61

***p≤.01
**p≤.05
*p≤.10

Conclusion

Not withstanding these contributions, there are some necessary precautions to be taken in understanding these findings. First, with the exception of Rhode Island, where the assessment was used for case planning for all inmates, the study samples were rather small. Although combined samples were adequate (the present study amassed data on 403 cases for gender-neutral variables and 626 cases for gender-responsive variables), the partitioning into construction and revalidation samples resulted in smaller samples. This necessitated a boot-strap approach where scales were developed in a construction validation sample and retested in a revalidation sample as well as in state-specific samples.

Second, one of the goals of this study was to refine the assessment in ways that would sharpen its predictive validity. To do so, additional questions were added to the interview and tested during the present analysis. In spite of training interviewers to the necessity of asking these test items, several did not. As a result, the missing data problem required median replacements for as many as 30 percent of the cases on a few items. This appeared to have attenuated the validity of some scales. In fact, findings for participants who were asked these items were higher than for those who were not asked these important questions.

Third, dynamic risk/need scales formed on the basis of interviews conducted in prison could change considerably during the post-prison supervision time frame. Evidence of this possibility is seen in a comparison of the Kentucky findings to those for Ohio, Missouri, and Rhode Island. The Kentucky interviews were conducted in home communities, after prison release, while those for the other sites were conducted prior to release. Correlations for the Kentucky sample were much higher than those for the other sites.

Fourth, the follow-up time period for the present study was 12 rather than 24 months. The earlier studies (2004-2008) found more impressive results at 24 months than at 12, and that is a standard observation for both program evaluations and prediction studies. Limited base rates are known to attenuate findings, and longer follow-up periods improve base rates, which in turn tends to improve predictive validity coefficients.

Fifth, though not shown in these analyses, results varied considerably from interviewer to interviewer. There were, in other words, interviewer effects, especially on scales pertaining to abuse and relationships. This has resulted in changes to WRNA training protocols in order to further build interviews' skills in building rapport and trust.

Finally, in many tests, results for the LSI-R and the WRNA-T, combined, were not as favorable as those for the WRNA-T alone. That is when the WRNA-T was added to the LSI-R, overall validity was "pulled down" by the LSI-R rather than the other way around. As was explained earlier, however, LSI-R

data was extracted from correctional files and may have been somewhat dated. Additionally, interviewers for the WRNA assessments were trained immediately prior to data collection. In contrast a number of state officials observed that many of the LSI-R interviewers were due to be retrained. Dynamic assessments such as the WRNA and the LSI-R require careful monitoring of quality assurance; validity of either assessments is likely to diminish when quality assurance becomes lax.

Even with these limitations, results are in keeping with those typically seen at a 12 month follow-up. The study has succeed in producing a shorter assessment, which finds on re-examination, that more of the gender-responsive variables are predictive than those observed during the original 2004-2008 research. Most importantly, we have much more confidence in the stability of the assessment, because it now is seen to be predictive across several jurisdictions rather than in the one, earlier, construction validation study which sampled only 149 cases.

Revalidation of the Women's Risk Needs Assessment:

Pre-Release Results[4]

Final Report

August 2012

Patricia Van Voorhis, Ph.D.
University of Cincinnati

Ashley Bauman, M.S., M.B.A.
University of Cincinnati

Rachel Brushett, M.S.
University of Cincinnati

[4] This research was funded by the National Institute of Corrections under cooperative agreement 09M12GKB3. Points of view or opinions stated in this article are those of the author and do not necessarily represent the official position or policies of the United States Department of Justice or the National Institute of Corrections.

Acknowledgements

Clearly, the present report is not solely the work of its three authors. Behind the scenes were funders, leaders of correctional agencies, interviewers, trainers, UC research staff, and research and planning staff at four of the five research sites. All worked hard to bring this research to fruition.

First, as noted throughout the report, this study was funded through a cooperative agreement between the University of Cincinnati and the National Institute of Corrections. We are especially appreciative of Maureen Buell's patience and constructive support throughout this process.

Correctional leaders at all of the sites provided access, vision and a good deal of support to our efforts. These included A.T Wall, Director of the Rhode Island Department of Corrections (RIDOC); Roberta Richman, Assistant Director (RIDOC): Carole Dwyer, Warden (RIDOC); Al Leach, Deputy Warden, (RIDOC); Terry Foley, Professional Services Coordinator (RIDOC); Ken Findley, Professional Services Coordinator (RIDOC); David Rost, Deputy Director of the Missouri Department of Corrections (MDOC); Tom Hodges, Chief State Supervisor of Probation and Parole (MDOC); Julie Kempker, Assistant Division Director of Probation and Parole (MDOC); Matt Sturm, Director of Offender Rehabilitation Services (MDOC); Marta Nolin, Assistant Division Director for Substance Abuse Services (MDOC); Jim Wiseman, Chief of Staff Training (MDOC); Tina Waldron, Former ReEntry Manager (MDOC); Cyndi Prudden, Deputy Director of Adult Institutions (MDOC); John Baron, Superintendent at the River City Correctional Center (RCCC); Helen Magers, Former Clinical Director (RCCC); Richard Schibi, Case Manager Coordinator (RCCC); Jodi Sleyo, Program Director (RCCC); Kimberly Sperber, Ph.D., Chief Research Officer at Talbert House; Kelly Hibner-Kalb, Program Manager at Rewards Jail Intervention (RJI); Maureen Kennedy, Former Clinical Supervisor (RJI); Brian Kowalski, Senior Researcher at the Ohio Department of Corrections and Rehabilitation (ODRC); John Chin, Chair of the Human Subjects Research Review Committee (ODRC); Steve Van Dine, Bureau of Research Chief (ODRC); and Staci Alder, Senior Case Manager at the Northern Kentucky Female Offender Re-Entry Project. In many cases, these leaders participated in planning sessions that contributed tremendous wisdom to our efforts.

Planning teams included the Women's Issues Committee of the Missouri Department of Corrections. Together with staff from the National Institute of Corrections and the University of Cincinnati, they authored the scales in 2003. Their contribution to this project was essential and the results speak highly of their expertise. A more contemporary team in Missouri, the Gender Responsive Assessment Implementation Team (GRAIT) was responsible for guiding the WRNA's implementation statewide.

The project was depended upon the assistance of staff in research and planning departments at two of the sites, Rhode Island and Missouri. We worked extensively with Bree Derrick and Erin Boyar at RIDOC and with David Oldfield and Fred Martin at MDOC. These individuals electronically compiled assessment and follow-up data for the project. We are grateful to them for the time, patience, and commitment they offered to this project.

This project could not have been completed without the dedication of interviewers and their trainers. My staff at the University of Cincinnati, including Ashley Bauman, Emily Wright, Krista Gehring, Valerie Bell and Rachel Brushett have logged many travel hours, prepared the training curriculum, and have effectively trained assessors in over 20 sites. The project could not have been completed without the dedication of correctional personnel who conducted interviews and in some cases provided case management to women offenders benefitting from the interviews. Their work serves as a model to others in the field. I can certainly thank the UC interviewers in this regard: Drs. Valerie Bell, Krista Gehring and William Stadler, as well as Ashley Bauman, Brittany Groot, Beth Ellefson, Noreen Loftus-Spilman, and Lindsay Morrow.

We are especially appreciative to the 600 plus women offenders who participated in this study. Their answers to assessment questions are the foundation of this work. Without the candid and honest disclosure of their life stories, this project would not have been possible. Moreover, many lend valuable input into the design of the assessment. They contributed their time willingly and without any remuneration. The future beneficiaries of this work owe them their gratitude.

Introduction

By the late 1990s, a number of scholars voiced concern for the applicability of the current generation of risk/needs assessments to women offenders. By then, dynamic risk/needs assessments had been widely adopted to address both security and treatment needs of correctional clientele. These assessment tools served the function of classifying offenders according to low, medium and high risk to assist agencies in managing the security needs of offenders. Additionally, they identified the needs or risk factors that were likely to contribute to offender recidivism. In doing so, these assessments also identified programmatic needs of offenders. Unfortunately, most of the widely used risk and need assessments were created for men and later applied to women prior to an examination of their appropriateness or validity (Bloom, Owen, & Covington, 2003; Chesney-Lind, 1997; Morash, Bynum, & Koons, 1998; Van Voorhis & Presser, 2001). Most importantly the assessments ignored needs central to women including: relationships, mental health problems, parental and childcare issues, safety, poverty, abuse and victimization, and strengths pertaining to family support, relationship support, self-efficacy, and educational attainments (Blanchette, 2004; Blanchette & Brown, 2006; Brennan, 1998; Brennan & Austin, 1997; Farr, 2000; Reisig, Holtfreter, & Morash, 2006; Van Voorhis, Wright, Salisbury & Bauman, 2010; and Van Voorhis, 2012).

To remedy this situation and other problems created by the lack of gender-responsive assessments, the National Institute of Corrections and the University of Cincinnati entered into a cooperative agreement to create and validate a women's, dynamic, risk/needs assessment, the Women's Risk Needs Assessment (WRNA). Development of two types of gender-responsive assessments began in 1999 with a pilot study in the Colorado Department of Corrections and later continued with three projects in Maui, Minnesota, and Missouri. The first, called the

Women's Risk Needs Assessment – Trailer (WRNA-T)(or "the trailer") was designed to supplement existing dynamic risk/needs assessments such as the Level of Service Inventory-Revised (LSI-R) (Andrews & Bonta, 1995) and the Northpointe COMPAS (Brennan, Dieterich, & Oliver, 2006) The second, the Women's Risk Needs Assessment (WRNA), was an assessment that could be used on its own, as a "stand-alone," dynamic, risk/needs assessment, comprised of both gender-neutral and gender-responsive scales. Extensive literature searches and focus groups with correctional administrators, treatment practitioners, line staff, and women offenders informed both of the assessments. Both instruments contained an interview and a self-report survey. The full instrument, and many of the questions now contained on the WRNA-T, was developed by members of the Women's Issues Committee of the Missouri Department of Corrections (MDOC) in collaboration with researchers at the University of Cincinnati and key staff from the National Institute of Corrections. This construction process also benefitted from the expertise of substance abuse specialists, psychologists and other mental health professionals on staff with MDOC.

The Women's Risk Needs Assessment (WRNA) was informed by two perspectives on offender rehabilitation: a) research by Canadian scholars Donald Andrews, Paul Gendreau, James Bonta, and others, which stressed the importance of assessing and treating dynamic risk factors (see Andrews & Bonta, 2010; Gendreau, Little & Goggin, 1996); and b) scholarship by feminist criminologists (e.g., Belknap, 2007; Bloom et al., 2003; Chesney-Lind, 1997; Daly, 1992; Morash, 2006; 2010) stressing the importance of women's unique "pathways" to crime. Both perspectives were relevant to the importance of programming targeted to dynamic risk factors for women offenders. However, the pathways perspective asserted that women's unique

needs were not adequately tapped by the current generation of risk/needs assessments, such as the LSI-R and the COMPAS.

The construction validation studies also produced different versions for specific types of correctional populations, because it was discovered that the predictive validity of both the gender neutral and the gender-responsive variables varied by correctional settings, e.g., prerelease, probation, and prisons.

In 2009, the National Institute of Corrections (NIC) entered into a second cooperative agreement with the University of Cincinnati, which produced the present study. Since the earlier assessments were created through construction validation, a key goal of the present study was to revalidate the original versions on new samples of offenders to assess the level of shrinkage in predictive validity from the construction to revalidation studies. Additionally, the 2009 cooperative agreement sought to refine several of the dynamic risk/needs scales in order to further improve predictive validity. This round of research tested a number of new items, listed on the assessment as "case management questions", that allowed for the exploration of their potential contributions to a revised assessment. Of course, creating a revised assessment also required another revalidation. To accomplish that, the new studies furnished larger samples than produced by the construction validation research and afforded an opportunity to partition the combined samples into construction and revalidation samples.

The present report focuses on the Pre-Release WRNA, the assessment intended for use in reentry planning for women released from prison. The rationale for seeking to improve the Pre-Release WRNA was attributable to the fact that original tool was developed on a small Missouri sample (N=149). The present study secured a sample of 403 participants across three sites, Missouri, Ohio, and Kentucky to test gender-neutral variables and 626 participants across 4 sites,

Missouri, Ohio, Kentucky, and Rhode Island, to test gender-responsive variables. The specific goals of this portion of the scale revisions were as follows:

1. To test the contributions of new items to the predictive validity of specific risk/needs scales as well as to the total risk scale representing the cumulation of risk/need factors predictive of offense-related outcomes.

2. To assure that those scales were valid on samples that were not part of the construction of the new scale. In other words to revalidate the revised scale through a split half validity test.

3. To produce an assessment that was more likely to work across samples and not be sample specific. Up to this point, it has been necessary to validate the WRNA on specific samples as data became available (see Van Voorhis et. al, 2010). While it has been advantageous to jurisdictions have an assessment specifically tailored to their use, the process resulted in slightly different total scales for each sample. The intent of the present study was to develop a single, more universal, assessment that would be applicable across settings.

4. To develop a trailer for the LSI-R. The 2004-2008 construction validation study did not finalize a supplement to the LSI-R. It is possible to do so in the present study.

Description of Participating Jurisdictions

Missouri

Under the previous cooperative agreement (2004-2008), the Missouri Department of Corrections (MDOC) collaborated with UC and NIC in the creation of the WRNA. In the spring of 2010 the state began a pilot project to implement the assessment. Staff participated in a 3 day training conducted by UC trainers. Select institutional parole officers and caseworkers at both Chillicothe Correctional Center (CCC) and Women's Eastern Reception, Diagnostic, and Correctional Center (WERDCC – also known as Vandalia) administered the stand alone, pre-release WRNA to all female offenders within 30 days of their release to the community. For this

study, assessment data were collected over 5 months, netting a total of 195 cases.[5] De-identified data was transmitted to University researchers allowing for a 100% response rate.

Kentucky

The Kentucky Department of Corrections operates a small, grant-funded re-entry project in the Northern Kentucky area. In this project, women being released from the Kentucky Department of Corrections were able to participate in a re-entry program which assisted them by providing services and treatment. Only women being released to this specific geographic area were eligible to participate. Women who elected to participate in the re-entry program were given the pre-release, stand alone WRNA and results were used to determine the intensity of services and supervision (from the risk score) and nature of services provided (from need domain scores). The Kentucky Department of Corrections contributed 36 cases to this study.[6] No women refused to participate resulting in a 100% response rate.

Ohio

Data for this research-only sample were collected at three Community Based Correctional Facilities (CBCFs) in the greater Cincinnati area over a period of several months utilizing the pre-release standalone version of the WRNA. The first site, Rewards Jail Intervention (RJI), was a residential, jail-based chemical dependency program for women. This program offered education and discharge planning for clients related to chemical dependency,

[5] The Missouri study was reviewed and approved (#10122703) by the Institutional Review Board (IRB) at the University of Cincinnati in February 2011. Re-approval was granted in February 2012.

[6] IRB approval for this study (#10092001) was obtained in September 2010. Because it was classified as secondary data analysis, re-approval was not required.

mental health, prostitution, and antisocial attitudes. However, staff reported that they often served as overflow housing for the jail as well. Only post-conviction clients were allowed to participate in the study, and pre-trial women were excluded. The second facility, River City Correctional Center (RCCC), consisted of a six-month, residential treatment facility for female offenders, most of whom were substance-abusing offenders. The clientele at RCCC was varied with some clients between prison and living in the community while others were sentenced directly to RCCC. Programming related to many criminogenic needs such as antisocial attitudes, chemical dependency, and employment. The third site, Pathways for Women, was a residential halfway house and treatment facility for women. The average length of stay at Pathways for Women was 90 days or less. The Pathways facility provided programming for chemical dependency and criminal attitudes as well as some employment services. Both RCCC and Pathways allowed women to engage in a work release program. In all three facilities, researchers obtained a population list and proceeded to ask each client in a private room if she would be interested in participating in the study. The population list was updated as needed during the study time frame. In some cases, women did not report to the room when called by the administration. Researchers noted in some instances that the women had already heard about the study and their failure to meet with the researcher was a refusal to participate. In other instances it appeared the women could not be located (due to court appearances, treatment programs, work release, etc.). In some cases the reason for the absence was not clear. Due to this difficulty locating residents, it was not possible to determine with certainty the number of potential participants who declined to participate. In total, 172 women agreed to participate in the study, and there were 21 known refusals for a total response rate of 88% for this sample. LSI-R's were completed as part of standard procedure at two of the facilities in this group. Thus, for 137 of

those women, LSI-R's were obtained from institutional files, allowing for the analysis of these women in both the stand-alone and trailer samples. [7]

Rhode Island

The Rhode Island Department of Corrections houses female clients at all custody levels and also acts as county jail for most of Rhode Island. On average, women in this institution served sentences of 9 months. The RIDOC began conducting the pre-release WRNA-T assessment as a supplement to the LSI-R in spring of 2008 for all women in the facility who had received sentences of 6 months or longer. RIDOC utilized the results for case planning purposes. Case managers asked women within the facility who had an LSI-R and a WRNA-T in their files if they would consent to release their assessment and file data to the University for the purposes of this study. The case managers reported a 100% response rate. A total of 223 assessments were provided by RIDOC for the study.[8]

Two of these sites, Missouri and Rhode Island received site specific reports prior to the preparation of the present study (see, Van Voorhis, Brushett, & Bauman, 2012; Van Voorhis, Bauman, & Brushett, 2012). Ohio and Kentucky participated as research sites and therefore are receiving only the present report.

Methodology

A total of 626 women offenders participated in the pre-release study. Of those, 403 completed the WRNA in Missouri (N=195), Kentucky (N=36), and Ohio (N=172) and 360

[7] IRB approval was granted for this study (#08081201 and #10020404) in September 2008 and March 2010. Re-approvals have been granted annually.

[8] IRB approval for the Rhode Island study (#09120704) was granted in January 2010. Re-approval was granted in January 2011 and 2012.

completed the WRNA-T in Rhode Island (N=223). A subsample of the Ohio sample (N=137) was also used in the WRNA-T analysis.

Sample Description

Table 1 presents demographic and criminal history characteristics for the three stand-alone samples that were utilized for the examination of the WRNA stand-alone assessment. As can be seen, in Table 1, the samples were comparable with only minor variations among them.

The average ages of women in all of the samples ranged from 33.0 years of age (Kentucky and Ohio) to 36.0 years of age (Missouri). Racial distributions were similar across samples; the majority of women in all samples were White. However, higher percentages of African-American's were observed for the Missouri sample (20.6 percent) compared to the Kentucky sample (2.8 percent) and the Ohio sample (7.0 percent).

In all of the samples, less than a third of the participants were married at the time of the study, with women in the Kentucky sample being the least likely to be married (19.4 percent). Much larger percentages of each of the groups had minor children, including 55.4 percent of the Missouri sample, 72.2 percent of the Kentucky sample, and 62.2 percent of the Ohio sample. Across all samples, over half of the women assessed possessed a high school diploma or general equivalency diploma, with women in the Ohio sample being the most likely to have this educational attainment (71.5 percent). A majority of those in the Missouri and Kentucky samples were employed prior to their arrest (69.2 percent and 66.7 percent, respectively) while less than half of those in the Ohio sample reported employment prior to their arrest (48.8 percent).

Regardless of sample, drug-related offenses were the most common current offense followed by larceny and forgery/fraud. Violent offenses were more common in the Missouri sample (11.3 percent) compared to the Kentucky and Ohio samples (5.6 percent and 9.4 percent,

Table 1. Frequency and Percent Distribution of Demographic and Offense-Related Background Characteristics of Samples Participating in the Validation of the Stand-Alone WRNA.

	Missouri		Kentucky		Ohio	
Characteristic	**N**	**Percent**	**N**	**Percent**	**N**	**Percent**
	195	100.0	36	100.0	172	100.0
Age						
18-20 years old	2	1.0	1	2.8	7	4.1
21-30 years old	51	26.2	14	38.9	77	44.8
31-40 years old	79	40.5	13	36.1	48	27.9
41-50 years old	49	25.1	7	19.4	31	18.0
51 years and older	14	7.2	1	2.8	9	5.2
	$\bar{X}$ = 36.0 yrs		$\bar{X}$ = 33.0 yrs		$\bar{X}$ = 33.0 yrs	
Race	N = 194					
Asian	3	1.5	0	0.0	0	0.0
African American	40	20.6	1	2.8	12	7.0
Hispanic/Latina	0	0.0	0	0.0	1	0.6
Native American	1	0.5	0	0.0	1	0.6
Other	2	1.0	0	0.0	3	1.7
White	148	76.3	35	97.2	155	90.1
Currently Married	N = 190					
Yes	53	27.9	7	19.4	40	23.3
Client Have Children Under 18						
Yes	108	55.4	26	72.2	107	62.2
Employment						
Employed (full or part-time, child care, student, or disabled)	135	69.2	24	66.7	84	48.8
Not employed	60	30.8	12	33.3	88	51.2
H.S. Grad or GED						
Yes	133	68.2	24	66.7	123	71.5

Table Continues

Table 1. Frequency and Percent Distribution of Demographic and Offense-Related Background Characteristics of Samples Participating in the Validation of the Stand-Alone WRNA, continued

	Missouri		Kentucky		Ohio	
Characteristic	**N**	**Percent**	**N**	**Percent**	**N**	**Percent**
	195	100.0	36	100.0	172	100.0
Most Serious Present Offense	N = 194				N = 171	
Arson	4	2.1	0	0.0	0	0.0
Assault	10	5.2	1	2.8	6	3.5
Burglary	11	5.7	4	11.1	13	7.6
Damage property	0	0.0	0	0.0	0	0.0
Dangerous drugs	62	32.0	16	44.4	61	35.7
DWI	10	5.2	1	2.8	14	8.2
Family offenses	7	3.6	2	5.6	7	4.1
Forgery/Fraud	35	18.0	5	13.9	15	8.8
Homicide/Manslaughter	6	3.1	0	0.0	1	0.6
Kidnapping	0	0.0	0	0.0	0	0.0
Larceny	31	16.0	6	16.7	30	17.5
Other	8	4.1	1	2.8	7	4.1
Robbery	4	2.1	0	0.0	4	2.3
Sex offenses	1	0.5	0	0.0	6	3.5
Stolen property	1	0.5	0	0.0	6	3.5
Traffic offenses	1	0.5	0	0.0	1	0.6
Weapon offenses	3	1.5	0	0.0	0	0.0
Present Offense Violent	N = 194					
Yes	22	11.3	2	5.6	16	9.4
Prior Felonies	N = 194					
Yes	130	67.0	17	47.2	59	34.3
Prior Incarcerations	N = 194					
Yes	136	70.1	22	61.1	18	10.5

respectively). Women in Missouri and Kentucky had more extensive criminal backgrounds than women in the Ohio sample. This held true for both prior felonies (67.0 percent in Missouri, 47.2 percent in Kentucky, and 34.3 percent in Ohio) and prior incarcerations (70.1 percent in Missouri, 61.1 percent in Kentucky, and 10.5 percent in Ohio). The prior offense distinctions in the Ohio sample are not surprising, given that these participants were in community residential settings rather than state prisons.

Table 2 presents demographic and criminal history characteristics for the two WRNA-T WRNA samples. As can be seen, in Table 2, the samples were again similar with only minor variations between them.

The average age of women in the Rhode Island sample was slightly higher than that of the Ohio sample (35.3 years of age versus 33.2 years of age). The majority of women in both samples were White (68.0 percent in Rhode Island and 94.2 percent in Ohio). Higher percentages of African-American's were seen within the Rhode Island sample (17.5 percent) compared to the Ohio sample (3.6 percent).

Women in the Ohio sample were slightly more likely to be married than women in the Rhode Island sample (22.6 percent compared to 17.0 percent), and a majority in both samples had minor children (60.6 percent in Ohio and 51.6 percent in Rhode Island). Almost two-thirds of the women in Ohio (73.0 percent) possessed a high school diploma or general equivalency diploma compared to only 41.3 percent of women in the Rhode Island sample. Half of the women in each sample reported employment prior to the current offense.

Just as in the stand-alone samples, drug-related offenses were the most common current offense followed by larceny and forgery/fraud. Violent offenses were uncommon in both samples (17.9 percent in Rhode Island and 11.0 percent in Ohio). Again, women in Rhode Island

Table 2. Frequency and Percent Distribution of Demographic and Offense-Related Background Characteristics of Samples Participating in the Validation of the Supplemental (WRNA-T) WRNA.

	Rhode Island		Ohio	
Characteristic	**N**	**Percent**	**N**	**Percent**
	223	100.0	137	100.0
Age				
18-20 years old	10	4.7	4	2.9
21-30 years old	68	31.6	60	43.8
31-40 years old	69	32.1	41	29.9
41-50 years old	56	26.0	25	18.2
51 years and older	12	5.6	7	5.1
	$\overline{X}$ = 35.3 yrs		$\overline{X}$ = 33.2 yrs	
Race	N = 200			
Asian	0	0.0	0	0.0
African American	35	17.5	5	3.6
Hispanic/Latina	22	11.0	0	0.0
Native American	4	2.0	0	0.0
Other	3	1.5	3	2.2
White	136	68.0	129	94.2
Currently Married				
Yes	38	17.0	31	22.6
Client Have Children Under 18				
Yes	115	51.6	83	60.6
Employment				
Employed (full or part-time, child care, student, or disabled)	113	50.7	69	50.4
Not employed	110	49.3	68	49.6
H.S. Grad or GED				
Yes	92	41.3	100	73.0

Table Continues

Table 2. Frequency and Percent Distribution of Demographic and Offense-Related Background Characteristics of Samples Participating in the Validation of the Supplemental (WRNA-T) WRNA, continued

	Rhode Island		Ohio	
Characteristic	**N**	**Percent**	**N**	**Percent**
	223	100.0	137	100.0
Most Serious Present Offense	N = 202		N = 136	
Arson	1	0.5	0	0.0
Assault	19	9.4	5	3.7
Burglary	8	4.0	12	8.8
Damage property	1	0.5	0	0.0
Dangerous drugs	54	26.7	55	40.4
DWI	7	3.5	6	4.4
Family offenses	7	3.5	7	5.1
Forgery/Fraud	14	6.9	15	11.0
Homicide/Manslaughter	2	1.0	0	0.0
Kidnapping	0	0.0	0	0.0
Larceny	32	15.8	19	14.0
Other	18	8.9	4	2.9
Robbery	10	5.0	4	2.9
Sex offenses	22	10.9	3	2.2
Stolen property	1	0.5	6	4.4
Traffic offenses	5	2.5	0	0.0
Weapon offenses	1	0.5	0	0.0
Present Offense Violent	N = 201		N = 136	
Yes	36	17.9	15	11.0
Prior Felonies	N = 219			
Yes	79	36.1	46	33.6
Prior Incarcerations	N = 219			
Yes	130	59.4	9	6.6

prison had more extensive criminal histories than women in the Ohio community residential settings, both in terms of prior felonies (36.1 percent in Rhode Island and 33.6 percent in Ohio) and prior incarcerations (59.4 percent in Rhode Island and 6.6 percent in Ohio).

Analytic Process

As noted above, the goals of the present study involved validating of the original Pre-Release WRNA, and examining ways to improve separate risk/need domains and the total risk/needs scale. The final dynamic risk/needs scale, to be used for overall risk assessment, was the sum of individual risk/needs scales determined to be associated with post-release recidivism. Two such scales were examined, a stand-alone WRNA and a WRNA-T. The WRNA-T was designed as a supplement to gender-neutral risk assessments, such as the Northpointe COMPAS or the LSI-R, with gender-responsive scales. The present study tested a trailer for the LSI-R.[9] The following analytical steps were employed:

1. Individual risk/need scales developed for a small construction validation sample (N=149) were tested, through analysis of correlations and AUC values with outcome measures. These tests involved the same items and scoring protocols resulting from the 2004-2008 construction validation study. Analyses were run twice, once for the full assessment and once for the WRNA-T assessment.

2. The original total risk/needs score (developed through construction validation research), including risk levels, was tested on the research samples for the present study (Missouri, Ohio, and Kentucky). Additionally, the cumulative WRNA-T scales were added to the LSI-R for the Ohio and Rhode Island samples and tested for predictive validity.

3. The current study collected data on a number of new test items to determine whether they improved the predictive validity of individual domain/need scales. Items were tested on a split half sample of all prerelease sites studied. The total sample was divided into a construction sample (N=201 for gender-neutral scales and 314 for gender-responsive scales) and a revalidation sample (N=202 for gender-neutral scales and 312 for gender-responsive scales). The construction and revalidation samples were drawn through a systematic random selection process where a pool of all cases was available and every

[9] In order to avoid use of redundant scales, the composition of the WRNA Trailer was specific to the gender-neutral assessment being used. For example, the Northpointe COMPAS did not contain mental health scales. Therefore, the COMPAS Trailer includes all of the WRNA Mental Health Scales, Mental Health History, Depression/Anxiety, and Psychosis. In contrast, he LSI-R, has a global mental health scale, Emotional/Personal. Therefore the LSI-R Trailer included only the Depression/Anxiety Scale and the Psychosis Scale of the WRNA.

other case was selected for the construction sample, and the remaining cases reserved for the revalidation sample. Items were developed on the construction validation sample and retested (confirmed) on the revalidation sample. A description of each of the samples is shown on Table 3. Table 3 shows very similar distributions across samples, and no significant differences on any of the background data tested.

4. Because both the construction and revalidation samples were small, another analysis was conducted which tested the new scales for each state sample. Scales found to be predictive in the construction validation study but not in the revalidation study, could nevertheless be retained for the final assessment if they were found to be predictive in two or more of the state samples. This rather unusual procedure accommodates some degree of sample specificity which occurred for both the WRNA and the LSI-R risk/need scales. That is, there was a tendency for a given risk/need domain to be predictive in some samples and not in others, regardless of the assessment used (LSI-R or WRNA). Therefore, we endeavored to create an assessment that would be predictive across samples even if the most predictive items varied from sample to sample.

5. Selection of a final risk/needs stand-alone and WRNA-T scales considered both the results for the construction and revalidation samples (step 3) and the state- specific findings (step 4). Total risk/needs scales were developed in the construction validation sample, retested on the revalidation sample, and then tested for specific sites.

Data analysis employed bivariate correlations (Pearson's r) and analysis of receiver operating characteristics (AUC). Psychometric properties of the new scales involved factor analysis (principal component extraction and varimax rotation) and alpha reliability measures. Results for factor analysis are not shown in this report but are available from the lead author.

Offense-Related Outcome Measures

Most participants were followed-up for 12 months, and results were reported at a 6 month interval and for the entire 12 months. Because parolees could fail (recidivate) in a variety of ways, a number of outcome measures were examined: a) NEW ARRESTS (Y/N); 2) NEW CONVICTIONS (Y/N); 3) RETURNS TO PRISON (through technical or law violations or new

Table 3. Frequency and Percent Distribution of Demographic and Offense-Related Background Characteristics of the Construction and Revalidation Samples.

	Construction		Revalidation		Total	
Characteristic	**N**	**Percent**	**N**	**Percent**	**N**	**Percent**
Age	N=201		N=202		N=403	
18-20 years old	7	3.5	3	2.8	10	2.5
21-30 years old	48	23.9	55	38.9	103	26.5
31-40 years old	102	50.7	106	36.1	208	51.6
41-50 years old	34	16.9	31	19.4	65	16.1
51 years and older	10	5.0	7	2.8	17	4.2
	$\overline{X}$ = 34.5 yrs		$\overline{X}$ = 33.4 yrs		$\overline{X}$ = 33.9 yrs	
Race	N = 280		N=276		N=556	
Asian	2	0.7	1	0.4	3	0.5
African American	45	16.1	43	15.6	88	15.8
Hispanic/Latina	11	3.9	12	4.3	23	4.1
Native American	5	1.8	1	0.4	6	1.1
Other	1	0.4	7	2.5	8	1.4
White	201	71.8	199	72.1	400	71.2
Currently Married	N = 314		N=308		N=622	
Yes	74	23.6	65	21.1	139	22.3
Client Have Children Under 18	N=315		N=312		N=627	
Yes	176	55.9	180	57.7	356	56.8
Employment	N=315		N=312		N=627	
Employed (full or part-time, child care, student, or disabled)	178	56.5	178	57.1	356	56.8
Not employed	137	43.5	134	42.9	271	43.2
H.S. Grad or GED	N=202		N=202		N=404	
Yes	147	72.8	134	66.3	281	69.6

Table 3. Frequency and Percent Distribution of Demographic and Offense-Related Background Characteristics of the Construction and Revalidation Samples, continued.

	Construction		Revalidation		Total	
Characteristic	**N**	**Percent**	**N**	**Percent**	**N**	**Percent**
Present Offense Violent	N = 202		N=202		N=404	
Yes	33	16.3	30	14.9	63	15.6
Prior Felonies	N = 202		N=202		N=404	
Yes	96	47.5	111	55.0	207	51.2
Prior Incarcerations	N = 202		N=202		N=404	
Yes	58	28.7	62	30.7	120	29.7

arrests/convictions); 4) ANY OFFENSE-RELATED FAILURE (e.g., new arrests/convictions as well as behavior which could have been processed as a violation but through officer discretion or agency policy was nevertheless recorded as a violation), and 5) ANY FAILURE (any of the above). As can be seen from Table 4, there was considerable variation across sites as to how offenders failed. In Missouri, for example, offenders were more likely to receive violations than arrests and convictions. There are very few arrests and even fewer convictions in the Missouri follow-up data[10], and returns to prison were through a violation process. As such measures of arrests and convictions are not as meaningful for that sample as they were for Ohio, Kentucky, or Rhode Island. The various Kentucky follow-up measures appeared to be redundant. Finally, base rates on the RETURN TO PRISON measure were very low for the Ohio sample. It is likely that the summary measure of ANY OFFENSE-RELATED FAILURE is the most comparable across sites.

[10] In fact, no convictions were noted in the Missouri follow-up data during the first 6 months.

Table 4. Follow-up Measures by Time Frame and Site

Site	Arrests		Conv.		Ret Pris[a]		Offense Fail		Any Fail	
	N	%	N	%	N	%	N	%	N	%
6 Month Follow-up										
Missouri (N=195)	14	7.2	0	0.0	17	8.7	30	15.4	99	50.8
Kentucky (N=36)	8	22.2	4	11.1	8	22.2	9	25.0	9	25.0
Ohio (N=172)	40	23.3	21	12.2	12	7.0	51	29.7	62	36.0
Ohio-WRNA-T (N=137)	30	21.9	16	11.7	7	5.1	36	26.3	42	30.7
Rhode Island-WRNA-T (N=222)	44	19.7	24	10.8			50	22.4	98	43.9
12 Month Follow-up										
Missouri (N=187)	24	12.8	6	3.2	37	19.8	61	32.6	125	66.8
Kentucky (N=35)	14	40.0	7	20.0	14	40.0	15	42.9	16	45.7
Ohio (N=169)	57	33.7	32	18.9	23	13.9	71	42.0	85	50.3
Ohio-WRNA-T (N=134)	45	33.6	25	18.7	15	11.2	53	39.6	62	46.3
Rhode Island-WRNA-T (N=210)	69	32.9	39	18.6			76	36.2	121	57.6

[a] Data on returns to prison were not available for Rhode Island.

The transmission of follow-up data also varied by site. Missouri DOC research personnel downloaded electronic data from the agency's information system. Rhode Island research personnel obtained follow-up data from a variety of State information sources and transmitted data electronically. Returns to prison were not sought. Kentucky data were sent by parole and program supervisors who had working knowledge of the case. Finally, Ohio data collection involved UC research personnel examining county court web sites, where offender records were recorded. Officials reported that arrests, convictions and returns to prison were likely to be more accurate than the records for technical violations.

Results

Revalidation of Individual Risk/Need Scales Created During Construction Validation Research

Tests of the original WRNA scales developed in construction validation research between 2004 and 2008 are shown in Tables 5 through 8. These tests use data collect during the current cooperative agreement from the sites described above. Table 5 and Table 6 show findings for all of the original WRNA scales, where both the gender-neutral scales and the gender-responsive scales were created by the earlier NIC/UC research. For tables 7 and 8, the gender neutral scales consisting of criminal history, antisocial thinking, education/employment, financial, family marital, accommodations, leisure/recreation, antisocial history, substance abuse history, were obtained from the LSI-R that was currently in use in Rhode Island and Ohio.

As suggested earlier, results for the stand alone scales, shown in Tables 5 and 6, varied by sample and outcome measure. The measures that were predictive for all three sites at 6 months (on at least one of the outcome measures) were antisocial attitudes, housing safety, antisocial friends, child abuse, sexual abuse, current/recent substance abuse, and parental

difficulties. Other measures, such as criminal history, employment financial, anger, adult victimizations, physical abuse, parental stress, and self-efficacy were associated with outcomes in some samples but not others. By 12 months, only antisocial associates and current/recent substance abuse predicted in all settings. Results were also favorable for criminal history (Ohio), employment/financial (Missouri and Ohio), adult victimization (Kentucky and Ohio), substance abuse history (Kentucky), depression (Ohio), relationship satisfaction (Kentucky), relationship dysfunction (Kentucky and Missouri), family support (Kentucky and Missouri), and parental stress (Missouri and Ohio).

Tables 7 and 8 present gender-neutral results for the LSI-R (shaded rows) and gender-responsive results for the WRNA. These sites tested the trailer to the LSI-R. It can be seen that a number of the LSI-R scales were predictive on a sample specific basis as well. These included antisocial attitudes, accommodations, leisure/recreation at 6 months, and antisocial thinking (attitudes), at 6 months and financial, leisure recreation, and substance abuse history at 12 months. In these two sites (Ohio and Rhode Island) many of the gender-responsive variables, including employment/financial, housing safety, anger, psychosis, child abuse, parental involvement, self-efficacy, at 6 months and employment/financial, anger, depression, adult abuse, sex abuse, parental involvement at 12 months were predictive for both samples. Others scales tended to be predictive in one sample but not the other regardless of follow-up time frame. Family support, family conflict, and relationship scales often were not predictive.

Table 5. Bivariate Relationship between Original WRNA Assessment Scales and 6-Month Offense-Related Outcomes, by Site.

	Missouri-WRNA (N=195)				
	Arrests	**Conv.**	**Ret. Pris**	**Offense Fail**	**Any Fail**
Scale	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**
Criminal history					
Attitudes	.10*				.15**
Educational needs	-.17***			-.11*	.09*
Educational assets (strength)	.12**				
Employment/financial			.09*		.24***
Housing safety	.10*		.12**	.14**	
Antisocial friends			.18***		.16***
Anger					
Mental health history					
Depression (symptoms)			.10*		.10*
Psychosis (symptoms)					
Child abuse					.16**
Adult abuse					
Sex abuse (adult or child)					.09*
Physical abuse (adult or child)					.12*
Substance abuse history	-.11*		.12**		
Substance abuse (current)			.15**		.15**
PTSD[a]			.24***	.15*	
Parental difficulties			-.12	-.09*	.10*
Parental involvement[b]	-.12*				
Relation. satisfaction (strength)					
Family conflict					.11*
Family support (strength)					-.11*
Relationship dysfunction					.11*
Parental stress (all)					
Child abuse (sur.)				.12**	.13**
Adult abuse (sur.)					
Self-efficacy (strength)	.10*		-.15**		

[a] The scale has excessive missing data and could not be used..
[b] Scale pertains to mothers, only.
***p≤.01
**p≤.05
*p≤.10

Table 5. Bivariate Relationship between Original WRNA Assessment Scales and 6-Month Offense-Related Outcomes, by Site, Continued.

	Kentucky: WRNA (N=36)				
	Arrests	**Conv.**	**Ret. Pris**	**Offense Fail**	**Any Fail**
Scale	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**
Criminal history					
Attitudes	.34**		.34**	.29**	.29**
Educational needs					
Educational assets (strength)					
Employment/financial					
Housing safety		.31**	.31**	.33**	.33**
Antisocial friends	.25*	.22*	.25*	.30**	.30**
Anger					
Mental health history					
Depression (symptoms)					
Psychosis (symptoms)					
Child abuse	.35**	.32**	.35**	.39***	.39***
Adult abuse		.34**	.26*		
Sex abuse (adult or child)	.30**	.37***	.30**	.29**	.29**
Physical abuse (adult or child)	.30**	.37***	.38***	.35**	.35**
Substance abuse history	.28**		.26*	.28**	.28**
Substance abuse (current)				.31**	.31**
PTSD[a]		-.25*			
Parental difficulties	.23*			.24*	.24*
Parental involvement[b]					
Relation. satisfaction (strength)	-.25*		-.25*		
Family conflict	-.27*		-.27*	-.29**	-.29**
Family support (strength)	-.22*		-.22*	-.24*	-.24
Relationship dysfunction					
Parental stress (all)					
Child abuse (sur.)			.28**	.26*	.26*
Adult abuse (sur.)					
Self-efficacy (strength)					

[a] The scale has excessive missing data and could not be used.
[b] Scale pertains to mothers, only.
***p≤.01
**p≤.05
*p≤.10

Table 5. Bivariate Relationship between Original WRNA Assessment Scales and 6-Month Offense-Related Outcomes, by Site, Continued.

	Ohio: WRNA (N=172)				
	Arrests	**Conv.**	**Ret. Pris.**	**Any Fail**	**Any Fail**
Scale	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**
Criminal history	.13**	.18***	.16**	.21***	.24***
Attitudes		-.15**	.15**		
Educational needs					
Educational assets (strength)		-.13**		-.10*	-.11*
Employment/financial	.18***	.10*	-.11*	.13**	.11*
Housing safety	.12*	.12*			
Antisocial friends	.11*	.18**	-.18**		
Anger	.21***	.13**		.16***	.12*
Mental health history			-.11*		
Depression (symptoms)					
Psychosis (symptoms)					
Child abuse		.12*	-.10*		
Adult abuse				.11*	
Sex abuse (adult or child)	.13**	.10*		.13**	
Physical abuse (adult or child)					
Substance abuse history		.13**			
Substance abuse (current)	.10*	.11*		.10*	
PTSD[a]		-.16*			
Parental difficulties	.12*	.19***		.11*	
Parental involvement[b]				-.13*	-.20**
Relation. satisfaction (strength)					
Family conflict					
Family support (strength)					
Relationship dysfunction					
Parental stress (all)	.17***	.16**		.19***	.16**
Child abuse (sur.)					
Adult abuse (sur.)					
Self-efficacy (strength)	-.15**		.13*		

[a] The scale has excessive missing data and could not be used.
[b] Scale pertains to mothers, only.
***p≤.01
**p≤.05
*p≤.10

Table 6. Bivariate Relationship between Original WRNA Assessment Scales and 12-Month Offense-Related Outcomes by Site.

	Missouri-WRNA (N=187)				
	Arrests	**Conv.**	**Ret. Pris**	**Offense Fail**	**Any Fail**
Scale	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**
Criminal history		.19***	-.09*		
Attitudes	.11*	-.11*	-.15**		
Educational needs	-.17***				
Educational assets (strength)	.12*				-.11*
Employment/financial	-.14**		.11*		.24***
Housing safety					
Antisocial friends		.19***	.25***	.21***	.16***
Anger		.10*			
Mental health history			.10*		
Depression (symptoms)			.10*		.10*
Psychosis (symptoms)					
Child abuse		.09*			.12**
Adult abuse					
Sex abuse (adult or child)					
Physical abuse (adult or child)					
Substance abuse history					
Substance abuse (current)	.10*		.20***	.29***	.16**
PTSD[a]		.22*		.15*	
Parental difficulties				.10*	.14**
Parental involvement[b]					
Relation. satisfaction (strength)					
Family conflict					.11*
Family support (strength)		-.17***			-.14**
Relationship dysfunction		.13**			.11*
Parental stress (all)	-.09*		-.12**	-.15**	
Child abuse (sur.)	.10*	.12**		.11*	
Adult abuse (sur.)					
Self-efficacy (strength)	.09*		-.11*		

[a] The scale has excessive missing data and could not be used.
[b] Scale pertains to mothers, only.
***p≤.01
**p≤.05
*p≤.10

Table 6. Bivariate Relationship between Original WRNA Assessment Scales and 12-Month Offense-Related Outcomes by Site, Continued.

	Kentucky: WRNA (N=35)				
	Arrests	**Conv.**	**Ret. Pris**	**Offense Fail**	**Any Fail**
Scale	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**
Criminal history					
Attitudes					
Educational needs					
Educational assets (strength)					
Employment/financial					
Housing safety		.25*	.27*	.30**	.24*
Antisocial friends	.42***	.38***	.42***	.47***	.43***
Anger					
Mental health history					
Depression (symptoms)					
Psychosis (symptoms)					
Child abuse					
Adult abuse		.34**	.26*		
Sex abuse (adult or child)		.24*			
Physical abuse (adult or child)		.24*			
Substance abuse history	.42***		.40***	.42***	.46***
Substance abuse (current)	.27*	.33**	.27*	.37**	.36**
PTSD[a]		-.25*			
Parental difficulties					
Parental involvement[b]					
Relation. satisfaction (strength)	-.33**		-.33**	-.29**	
Family conflict		-.23*	-.27*	-.29**	-.29**
Family support (strength)	-.22*		-.22*	-.24*	-.26*
Relationship dysfunction	.33**		.37**	.33**	.30**
Parental stress (all)					
Child abuse (sur.)					
Adult abuse (sur.)					
Self-efficacy (strength)					

[a] The scale has excessive missing data and could not be used.
[b] Scale pertains to mothers, only.
***p≤.01
**p≤.05
*p≤.10

Table 6. Bivariate Relationship between Original WRNA Assessment Scales and 12-Month Offense-Related Outcomes by Site, Continued.

	Ohio: WRNA (N=169)				
	Arrests	**Conv.**	**Ret. Pris**	**Offense Fail**	**Any Fail**
Scale	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**
Criminal history	.12**	.19***	.11*	.18***	.21***
Attitudes		-.10*	.11*		
Educational needs					
Educational assets (strength)					
Employment/financial	.18**			.15**	.13**
Housing safety					
Antisocial friends	.20***	.17***		.11*	
Anger	.14**	.18***			
Mental health history			-.11*		
Depression (symptoms)	.13**			.14**	.14**
Psychosis (symptoms)				.10*	
Child abuse			-.13**		
Adult abuse			.13**	.11*	.11*
Sex abuse (adult or child)	.13**			.13**	.15**
Physical abuse (adult or child)					
Substance abuse history					
Substance abuse (current)	.17***			.19***	.13**
PTSD[a]		-.25**			
Parental difficulties	.13*	.17**		.15**	.13**
Parental involvement[b]					
Relation. satisfaction (strength)					
Family conflict					
Family support (strength)					
Relationship dysfunction					
Parental stress (all)					
Child abuse (sur.)					
Adult abuse (sur.)	.11*			.11*	.11*
Self-efficacy (strength)	-.13**				

[a] The scale has excessive missing data and could not be used.
[b] Scale pertains to mothers, only.
***p≤.01
**p≤.05
*p≤.10

The results likely reflect a number of differences across jurisdictions, including base rates on follow-up measures, interviewer skill sets, programming resources, and cultural, ethnic and other offender background characteristics. The fact that the results were especially strong for the Kentucky sample, especially on the gender responsive variables, reflects one of the challenges in administering dynamic risk/needs assessments to pre-release samples. Results for assessments administered in prisons are likely to change upon release to communities. Kentucky was the only pre-release settings where assessments were administered after release, after the offenders had returned to their home communities.

Another pattern in these findings, suggests that findings were better in those sites where the assessment was actually being used for treatment planning and case management. For example, Rhode Island and Kentucky evidenced especially strong results. The assessments administered in Ohio and Missouri were for research purposes. Missouri was experiencing problems with staff buy-in and Ohio assessments were administered by research personnel at the University of Cincinnati. Nevertheless, the Ohio results were more favorable than those for Missouri.

In considering the results for both the WRNA scales and the LSI-R scales, it appeared that sample variations would need to be accommodated. That is the overarching goal was for a predictive total scale with the realization that, in all likelihood, the total scale might be driven by different risk/need scales in different samples.

Table 7. Bivariate Relationship between LSI-R and Original WRNA Assessment Scales and 6-Month Offense-Related Outcomes, by Site.

	Ohio-LSI-R (N=137)					Rhode Island: LSI-R (N=223)				
	Arrests	Conv.	Ret Pris	Offense Fail	Any Fail	Arrests	Conv.	Ret Pris[a]	Offense Fail	Any Fail
Scale	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**
Criminal history	.16**		.14**	.23****	.23**	.21***	.19***		.26***	.13**
Attitudes										.15**
Employment/Financial (WRNA)	.16**		-.11*	.12*	.13*	.13**			.14**	.12**
Education/employment										
Financial		.14**			.18**		.09*			.10*
Housing safety (WRNA)		.14**							.10*	
Family/Marital	.14**	.11*		.13*						.10*
Accommodations						.11**				.14**
Leisure/recreation						.13**			.09*	.14**
Antisocial friends	.14**	.14**		.13*						.14**
Anger (WRNA)	.16**			.17**	.14*	.11**	.29***		.23***	.19***
Mental health history			.16**	.13*	.12*					.10*
Depression (symptoms) (WRNA)										
Psychosis (symptoms)(WRNA)			.12*		.12*	.09*			.10*	
Child abuse (WRNA)		.14**	-.11*							.10*
Adult abuse (WRNA)			.11*	.13*						
Sex abuse (adult or child) (WRNA)										.11*
Physical abuse (adult or child) (WRNA)										
Substance abuse history										
PTSD[b] (WRNA)										
Parental difficulties (WRNA)										
Parental involvement[c] (WRNA)					-.15**	-.20**			-.21***	
Relation. satisfaction (strength)(WRNA)						.13**			.10*	.13**

Table 7. Bivariate Relationship between LSI-R and Original WRNA Assessment Scales and 6-Month Offense-Related Outcomes, by Site, Continued.

	Ohio-LSI-R (N=137)					Rhode Island: LSI-R (N=223)				
	Arrests	Conv.	Ret Pris	Offense Fail	Any Fail	Arrests	Conv.	Ret Pris[a]	Offense Fail	Any Fail
Scale	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**
Family conflict (WRNA)										.10*
Family support (strength)(WRNA)										
Relationship dysfunction(WRNA)										
Child abuse (sur.) (WRNA)	.12*	.11*		.14*	.14**					
Adult abuse (sur.)(WRNA)						.09*	.15**		.11**	
Self-efficacy (strength)(WRNA)	-.14**					-.09*	-.11**		-.12**	-.15***
Parental stress(WRNA)	.16**	.19***		.18**	.17**					

[a] Data pertaining to returns to prison were not available.
[b] The scale had excessive missing data and could not be used.
[c] Scale pertains to mothers, only.
***p≤.01
**p≤.05
*p≤.10

Table 8. Bivariate Relationship between LSI-R and Original WRNA Assessment Scales and 12-Month Offense-Related Outcomes, by Site.

	Ohio-LSI-R (N=134)					Rhode Island: LSI-R (N=210)				
	Arrests	Conv.	Ret. Pris	Offense Fail	Any Fail	Arrests	Conv.	Ret. Pris[a]	Offense Fail	Any Fail
Scale	Y/N		Y/N	Y/N		Y/N	Y/N	Y/N	Y/N	Y/N
Criminal history	.23***			.26***	.25***	.29***	.24***		.33***	.21**
Attitudes			.12*						.09*	.13****
Employment/Financial (WRNA)	.18**			.18**	.15**	.21***	.13**		.20***	.16***
Education/employment			.11*	.18**						.11*
Financial							.11**		.12**	.17***
Housing safety (WRNA)							.17***		.10*	
Family/Marital			.13*			.09*	.11*		.10*	.10*
Accommodations							.09*			.09*
Leisure/recreation							.12**			.11*
Antisocial friends	.15**	.13*		.18**	.19**	.09*	.13**		.13**	.15**
Anger (WRNA)		.13*				.18***	.28***		.28***	.21***
Mental health history				.14**	.11*				.10*	.10*
Depression (symptoms) (WRNA)					.12*		.16***		.12**	
Psychosis (symptoms)(WRNA)		-.15**				.13**	.15**		.21***	
Child abuse (WRNA)			-.15**							.13**
Adult abuse (WRNA)			.16**				.14**			
Sex abuse (adult or child) (WRNA)					.16**	.11*			.09*	.14**
Physical abuse (adult or child) (WRNA)						.11*				
Substance abuse history										.14**
PTSD[a] (WRNA)	-.21	-.26*								
Parental difficulties (WRNA)										
Parental involvement[b] (WRNA)			-.16*			-.17**			-.14***	
Relation. satisfaction (strength)(WRNA)						.14**			.12**	.14**

Table 8. Bivariate Relationship between LSI-R and Original WRNA Assessment Scales and 12-Month Offense-Related Outcomes, by Site, Continued.

	Ohio-LSI-R (N=134)					Rhode Island: LSI-R (N=210)				
	Arrests	Conv.	Ret. Pris	Offense Fail	Any Fail	Arrests	Conv.	Ret. Pris[a]	Offense Fail	Any Fail
Scale	**Y/N**		**Y/N**	**Y/N**		**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**
Family conflict (WRNA)										.10*
Family support (strength) (WRNA)							-.13**			
Relationship dysfunction (WRNA)										
Child abuse (sur.) (WRNA)										
Adult abuse (sur.) (WRNA)						.13*	.17***		.13**	
Self-efficacy (strength) (WRNA)							-.16***			-.10*
Parental stress (WRNA)	.12*	.22***	.13*	.20***	.16**					

[a] Data pertaining to returns to prison were not available.
[b] The scale has excessive missing data.
[c] Scale pertains to mothers, only.
***p≤.01
**p≤.05
*p≤.10

Revalidation of the Original Cumulative Risk Scales Created During Construction Validation Research

In the 2004 to 2008 Missouri construction validation study, a cumulative scale was created from individual risk/needs scales shown in Appendix A. Scoring guidelines for that assessment are shown in Appendix B. Table 9 shows the results of the revalidation of the stand-alone scale in samples where the full WRNA was tested. Typically, our earlier studies and presentations have set a standard for a correlation of at least .27 and AUC values above .70 in order to conclude that an assessment was valid. This standard, however, was pertinent to samples with 24 months of follow-up. Using past research as a guide, predictive validity tends to be most strong at the 24 month point, when base rates are most favorable to the statistical tests. The present study provides only 12 months of follow-up.

Just the same, the Kentucky site achieved correlations and AUC values that were in a favorable range. At 6 months correlations ranged from .25 to .35 across outcome measures, and from .27 to .38 for the 12 month outcomes. Findings for Missouri and Ohio were much less impressive and in need of improvement. Nevertheless, they were typical of results seen at 12 months in the earlier research.

Table 9 also shows findings from a second analysis which was conducted on only the gender responsive cumulative scales, scales that would be used in a WRNA-T if a jurisdiction chose not to use the stand-alone assessment. These must be interpreted with caution, because this cumulative gender-responsive scale is not intended to be used as a full instrument as some have suggested (Andrews & Bonta, 2010). The point here is to show that as a block of factors, these scales are significantly related to outcomes. However, the scale is not intended to reach the strengths of association seen with full scales (gender-neutral plus gender-responsive). Table 9 notes in fine print several 6 and 12 month findings for Kentucky which did not reach significance

due to the small sample size and limited power of the statistical tests. Though not significant, the strength of these findings was similar to those for Ohio and Missouri.

Results for settings testing the WRNA-T and adding it to LSI-R cumulative scales are shown in Table 10, below. Again, these use the scales developed during the construction validation study conducted between 2004 and 2008. Note that the Rhode Island study did not furnish data on RETURNS TO PRISON. For clarity of presentation, we focus on the 6 and 12 month outcomes pertaining to OFFENSE-RELATED FAILURES and ANY FAILURE. At the 6 month follow-up point, the WRNA-T made a contribution to the overall predictive validity of the findings for Ohio. Correlations improved from $r=.13, p \leq .10$ *to* $r=.18, p \leq .05$ for offense related failures and from $r=.11, p \leq .10$ *to* $r=.16, p \leq .05$ for any failure. Additionally, the WRNA-T results by themselves were more favorable than those for the LSI-R. The contribution of the WRNA-T items (partial correlation) was statistically significant. Similar findings were observed for the Rhode Island sample at 6 months.

By 12 months, however, the predictive validity of the LSI-R improved for the Ohio sample. While the relationship between the WRNA-T and the two outcomes measures was significant in Ohio and made a very modest improvement to the LSI-R, the partial correlation was not significant. This was not the case for Rhode Island where across all outcome measures, the WRNA-T made an important contribution to the predictive validity of the LSI-R.

Table 9. Bivariate Relationship between Original WRNA Assessment Scales and 6 and 12 Month Offense-Related Outcomes, by Site.

	Arrests		Conviction		Return to Prison		Offense-Related Failure		Any Failure	
Scale	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**
6 Month Follow-up for Original Stand-Alone WRNA										
Missouri Original WRNA (N=195)					.15**	.64			.20***	.62
Missouri-Levels					.13**	.60	.12**	.57	.25***	.62
Kentucky Original WRNA (N=36)	.35**	.78			.25*	.71	.32***	.75	.32***	.75
Kentucky-Levels	.33**	.70	.22*	.67			.26*	.65	.26*	.65
Ohio Original WRNA (N=172)	.21***	.64	.19***	.63	-.10*		.16**	.60	.14**	.58
Ohio-Levels	.18***	.61					.15**	.58	.12*	.56
6 Month Follow-up Original WRNA-T Scales										
Missouri Original WRNA-T (N=195)					.12**	.61	.10*	.55	.21***	.62
Kentucky Island Original WRNA-T (N=36)	.24**	.68	.16	.65	.12	.59	.19	.64	.19	.64
Ohio Original WRNA-T (N=172)	.25***	.66	.12*	.60			.20***	.62	.15**	.58
12 Month Follow-up for Original Stand-Alone WRNA										
Missouri Original WRNA (N=187)	-.10*		.18***	.76	.19***	.64	.15**	.59	.17***	.60
Missouri-Levels			.18***	.72	.18***	.61	.14**	.57	.14***	.57
Kentucky Original WRNA (N=35)	.38***	.73	.27*	.71	.29**	.68	.36**	.73	.31**	.69
Kentucky-Levels	.35***	.68	.29**	.88	.24*	.62	.30**	.65	.25*	.62
Ohio Original WRNA (N=169)	.23***	.63	.20***	.63			.20***	.61	.15**	.58
Ohio-Levels	.21***	.61	.17***	.61			.21***	.69	.13***	.56
12 Month Follow-up for Original WRNA-T Scales										
Missouri Original WRNA-T (N=187)	-.16***				.14**	.61			.18**	.61
Kentucky Original WRNA-T (N=35)	.13	.58	.28*	.70						
Ohio Original WRNA-T (N=169)	.21***	.62	.13**	.60			.19***	.61	.15**	.59

***p≤.01
**p≤.05
*p≤.10

Revision of the WRNA Scales

Improvement of the assessment scales was achieved through the development and testing of new items on a construction validation sample and then revalidating those items. Presumably, the revalidation of the new scales will reduce concerns for the need of another revalidation study. The number of cases available for testing depended upon whether or not the scale was a gender-neutral or a gender-responsive scale. Gender-neutral scales were only tested in the 3 sites that examined the stand- alone assessment. Tests of criminal history, attitudes, educational needs and educational assets, antisocial friends, and substance abuse history scales were only tested in Ohio, Missouri and Kentucky. Sample sizes for the construction sample and revalidation samples consisted of 201 participants at 6 months, for the construction sample and 202 for the revalidation sample at 12 months. At 12 months, tests of the gender neutral scales involved 194 participants in the construction sample and 197 in the revalidation sample. Gender responsive scales were tested at all sites, and included 313 participants and 312 participants in the construction and validation samples, respectively, at 6 months. At 12 months, the samples consisted of 303 in the construction sample, and 298 in the revalidation sample.

Table 10. Bivariate Relationship between Original WRNA-T Assessment Scales and 6 and 12 Month Offense-Related Outcomes, by LSI-R Sites.

	Arrests		Conviction		Return to Prison		Offense-Related Failure		Any Failure	
Scale	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**
6 Month Follow-up										
Ohio LSI-R (N=137)							.13*	.58	.11*	.57
Ohio Original WRNA-T (N=137)	.21***	.64					.20***	.63	.17**	.61
Ohio LSI-R + WRNA-T(N=137)	.17**	.60					.18**	.61	.16**	.59
Partial corr	.19**						.17**		.15**	
Rhode Island LSI-R (N=223)	.11*	.57	.10*	.58			.10*	.56	.16***	.58
Rhode Island Original WRNA-T (N=223)	.16***	.61	.18***	.66			.22***	.64	.17***	.59
Rhode Island LSI-R + WRNA-T (N=223)	.13**	.57	.13**	.60			.15**	.58	.18***	.59
Partial corr.	.13**		.15***				.20***		.09*	
12 Month Follow-up										
Ohio LSI-R (N=134)	.16**	.59			.14**	.65	.23***	.64	.19***	.61
Ohio Original WRNA-T (N=134)	.13*	.58					.17**	.60	.14**	.58
Ohio LSI-R + WRNA-T (N=134)	.18**	.60			.12*	.64	.25***	.65	.20***	.62
Partial corr.										
Rhode Island LSI-R (N=210)	.14**	.57	.14**	.59			.18***	.60	.21***	.62
Rhode Island Original WRNA-T (N=210)	.26***	.65	.25***	.68			.29***	.67	.21***	.62
Rhode Island LSI-R + WRNA-T (N=210)	.19***	.59	.19***	.62			.23***	.62	.23***	.63
Partial corr.	.22***		.21***				.23***		.10*	

***p≤.01
**p≤.05
*p≤.10

These numbers are somewhat smaller than desired, so a second examination of revised scales was made in each of the research samples. Given the sample-specific nature of some findings, we retained scales that may not have reached adequate predictive validity in the revalidation sample, but did in at least two of the test sites. Results for the construction and revalidation samples are shown in Tables 11 and 12. Findings for the revised scale are compared to the original WRNA scales (2004-2008). State specific findings are shown in Appendix C for the stand-alone WRNA and for the WRNA-T. LSI-R domain scales are not the subject of these analyses. Alpha measures of internal consistency are shown in Appendix D. Discussion of changes made to specific scales follows.

CRIMINAL HISTORY: Problems with the original criminal history scale were known in advance, because findings were also observed to be rather weak during the construction validation study. It was assumed that the scale would be amended as part of the present revalidation study. To assist with this effort, research personnel in the Missouri Department of Corrections suggested six additional questions for the scale. The ones that contributed to the predictive validity of the scale were:

1. Was your last conviction within the past three years?
2. Age at intake
   18-30=2
   31-40=1
   40+=0

In addition to these changes validity was improved by omitting an item pertaining to misconducts during the current prison term. As shown in Table 11 and 12, these changes improved the scale considerably for construction and revalidation samples as well as for specific sites, where the scale was predictive of misconducts at all of the three sites. Even so, results are likely to be attenuated by excessive missing data on the two test variables, age (24.8 percent missing) and last conviction (24.8 percent missing). It was necessary to replace missing values at the median to retain this very important scale, but that clearly involved a higher number of replacements than desired.

The revised assessment scale also mandates use of official records of prior offenses. In most sites, this study had access to both self-report and official accounts. Discrepancies were detected, and the official records proved to be more valid and were substituted for self-report wherever possible. Alpha for the scale was low (.11) but that is typical for criminal history scales where items are intentionally dissimilar. The revised criminal history scale was included in the final stand-alone WRNA.

Table 11. Bivariate Relationship between Revised WRNA Assessment Scales and 6-Month Offense-Related Outcomes, by Construction and Revalidation Samples.

	Original WRNA: Full Sample				
	Arrests[a]	**Conv.[a]**	**Ret Pris[b]**	**Offense Fail**	**Any Fail**
Scale	**Y/N**		**Y/N**	**Y/N**	**Y/N**
Criminal history	.07*	.16***			.17***
Attitudes		-.10*			.11***
Educational needs					
Educational assets (strength)		-.09*			-.08*
Employment/financial	.14***	.08**		.13***	.14***
Housing safety	.11**	.12***	.08**	.13***	
Antisocial friends	.08*	.16***		.08**	..08*
Anger	.14***	.21***		.17***	.10***
Mental health history					
Depression	.05*			.07*	.07*
Depression (col.)[c]					
Psychosis (symptoms)	.06*			.07*	
Child abuse		.06*			.09*
Adult abuse				.06*	
Sex abuse (adult or child)	.07**			.07**	.10**
Physical abuse (adult or child)		.07*	.08**		
Substance abuse history		.12**	.06*		
Substance abuse (current)	.19***	.10*		.18***	
PTSD[a]	.13*				.11**
Parental involvement[b]	-.13**		-.08*	-.15***	
Relation. satisfaction (strength)					
Family conflict					
Family support					
Family support (col.)					
Relationship dysfunction					
Parental stress (all)		.07**			
Parental stress, recoded					
Child abuse (sur.)			.09**	.05*	.07*
Child abuse_col					
Adult abuse (sur.)		.11**		.06*	
Adult abuse (col.)					
Self-efficacy	-.11***	-.08*		-.11***	-.05*
Self-efficacy (col.)	-.12***		-.07**	-.12***	-.06

[a] Missouri cases are omitted from 6 mo. convictions.
[b] Rhode Island cases are omitted from return to prison analysis. Data were not available.
[c] Shaded areas indicate that the measure was not tested.
***p≤.01
**p≤.05
*p≤.10

Table 11. Bivariate Relationship between Revised WRNA Assessment Scales and 6-Month Offense-Related Outcomes, by Construction and Revalidation Samples, Continued.

	Revised WRNA: Construction				
	Arrests	**Conv.[a]**	**Ret Pris[b]**	**Offense Fail**	**Any Fail**
Scale	**Y/N**		**Y/N**	**Y/N**	**Y/N**
Criminal history		.18**	.14**	.11*	.27***
Attitudes		-.18**			.09*
Educational needs[c]					
Educational assets (strength)		-.14**			-.13**
Employment/financial	.17**	.16***		.13**	.07*
Housing safety	.10**	.25***	.18***	.17***	.15***
Antisocial friends					
Anger	.11**	.25***		.18***	.12***
Mental health history					
Depression	.09*	.19***		.11**	.11**
Depression (col.)	.10**	.14***		.10**	.11**
Psychosis (symptoms)					
Child abuse					
Adult abuse					
Sex abuse (adult or child)					
Physical abuse (adult or child)					
Substance abuse history					
Substance abuse (current)	.28***	.22**		.24***	
PTSD[a]					
Parental involvement[b]	-.20***	-.23***		-.24***	
Relation. satisfaction (strength)					-.10**
Family conflict					.09*
Family support		-.10*			-.09*
Family support (col.)		-.12**			
Relationship dysfunction		.18***	-.11*		.11**
Parental stress (all)					
Parental stress, recoded	.08*	.10**		.11**	
Child abuse (sur.)			.19***		.12**
Child abuse_col			.15**		.11**
Adult abuse (sur.)		.10**		.08*	.08*
Adult abuse (col.)		.09*	.09*		
Self-efficacy	-.13***	-.16***		-.10**	
Self-efficacy (col.)	-.13**	-.10**		-.08*	

[a] Missouri cases are omitted from 6 mo. convictions.
[b] Rhode Island cases are omitted from return to prison analysis. Data were not available.
[c] Shaded areas indicate that the measure was not tested.
***p≤.01
**p≤.05
*p≤.10

Table 11. Bivariate Relationship between Revised WRNA Assessment Scales and 6-Month Offense-Related Outcomes, by Construction and Revalidation Samples, Continued.

	Revised WRNA: Revalidation Sample				
	Arrests	**Conv.[a]**	**Ret. Pris[b]**	**Offense Fail**	**Any Fail**
Scale	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**
Criminal history		.21**	.24***	.20***	.29***
Attitudes	.19***			.17***	.14***
Educational needs[c]					
Educational assets (strength)					
Employment/financial	.09***	.11*		.08*	.13**
Housing safety	.20***			.18***	.14***
Antisocial friends					
Anger	.21***	.19**		.18***	.15***
Mental health history					
Depression					
Depression (col.)	.08*			.09**	
Psychosis (symptoms)					
Child abuse					
Adult abuse					
Sex abuse (adult or child)					
Physical abuse (adult or child)					
Substance abuse history		.16**		.10*	
Substance abuse (current)	.11*			.15*	
PTSD[a]					
Parental involvement[b]					
Relation. satisfaction (strength)					.09*
Family conflict	.12**		.09*	.16***	.15*
Family support					
Family support (col.)					
Relationship dysfunction					
Parental stress (all)	.10**	.11**		.11**	.17***
Parental stress, recoded	.13***			.13***	.16***
Child abuse (sur.)	.11**			.07*	
Child abuse_col					
Adult abuse (sur.)					
Adult abuse (col.)					
Self-efficacy	-.10**	-.09*	.10*	-.12**	-.09*
Self-efficacy (col.)	-.12**	-.11*	-.12**	-.14***	-.10**

[a] Missouri cases are omitted from 6 mo. convictions.
[b] Rhode Island cases are omitted from return to prison analysis. Data were not available.
[c] Shaded areas indicate that the measure was not tested.
***p≤.01
**p≤.05
*p≤.10

Table 12. Bivariate Relationship between Revised WRNA Assessment Scales and 12-Month Offense-Related Failures, by Construction and Revalidation Samples.

	Original WRNA: Full Sample				
	Arrests	**Conv.**	**Ret Pris[a]**	**Offense Fail**	**Any Fail**
Scale	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**
Criminal history					.15***
Attitudes			-.07*	.07*	
Educational needs					
Educational assets (strength)			-.08*		-.08
Employment/financial	.13***	.12***		.13***	.14***
Housing safety	.08**	.13***		.05*	
Antisocial friends	.19***	.21***	.12***	.20***	.10**
Anger	.13***	.22***		.15***	.07***
Mental health history				.06*	.06*
Depression (symptoms)	.10***	.12***		.11***	.08**
Depression (col.)[b]					
Psychosis (symptoms)	.08**	.07**		.12***	
Child abuse	.06*				.10**
Adult abuse		.07**	.09**	.06*	
Sex abuse (adult or child)	.07**	.07*		.07**	.11***
Physical abuse (adult or child)		.05*	.07**		
Substance abuse history	.11**	.11**	.07*	.13***	
Substance abuse (current)	.25***	.20***		.23***	
PTSD[a]					
Parental difficulties				.10** *	.07**
Parental involvement[b]	-.11**	-.10*		-.09*	
Relation. satisfaction (strength)				-.06*	
Family conflict					
Family support (strength)		-.06*			
Family support (strength)(col.)		-.09*			
Relationship dysfunction		.07*			
Parental stress (all)		.05*		.06*	.07*
Parental stress (col).					
Child abuse (sur.)				.06*	
Child abuse (col)					
Adult abuse (sur.)		.08*	.07*	.08*	
Adult Abuse (col)					
Self-efficacy (strength)	-.10***	-.13***		-.08*	
Self-efficacy (col.)	-.11***	-.10***		-.11***	

[a] Rhode Island cases are omitted from return to prison analysis. Data were not available.
[b] Shaded areas indicate that the measure was not tested.
***p≤.01
**p≤.05
*p≤.10

Table 12. Bivariate Relationship between Revised WRNA Assessment Scales and 12-Month Offense-Related Failures, by Construction and Revalidation Samples, Continued.

	Revised WRNA: Construction Sample				
	Arrests	**Conv.**	**Ret Pris[a]**	**Offense Fail**	**Any Fail**
Scale	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**
Criminal history			.26***	.21****	.25***
Attitudes		-.10*			.14**
Educational needs[b]					
Educational assets (strength)	-.13**		-.12*	-.15**	-.12**
Employment/financial	.15***	.14***		.12**	
Housing safety		.19***	.10*	.09*	.08*
Antisocial friends					
Anger	.11**	.25***		.12**	.09*
Mental health history					
Depression (symptoms)	.19***	.17***		.17***	.14***
Depression (col.)	.16***	.14***		.16***	.12**
Psychosis (symptoms)					
Child abuse					
Adult abuse					
Sex abuse (adult or child)					
Physical abuse (adult or child)					
Substance abuse history	.13**	.12**	.11*	.16***	.12**
Substance abuse (current)	.33***	.19***		.30**	.10*
PTSD[a]					
Parental difficulties					
Parental involvement[b]	-.19***	-.18***		-.17**	
Relation. satisfaction (strength)		-.08*			
Family conflict				.08*	
Family support (strength)		-.09*			
Family support (strength)(col.)		-.10**			
Relationship dysfunction	.09**	.13***		.09*	.12**
Parental stress (all)			.14**	.07*	
Parental stress (col).	.10**	.10**	.14**	.13***	
Child abuse (sur.)			.13**		.08*
Child abuse (col)			.13**		.08*
Adult abuse (sur.)		.09**	.13**	.10**	.09*
Adult Abuse (col)		.08*	.15**	.09**	
Self-efficacy (strength)					
Self-efficacy (col.)	-.16***	-.10**		-.10**	

[a] Rhode Island cases are omitted from return to prison analysis. Data were not available.
[b] Shaded areas indicate that the measure was not tested.
***p≤.01
**p≤.05
*p≤.10

Table 12. Bivariate Relationship between Revised WRNA Assessment Scales and 12-Month Offense-Related Failures, by Construction and Revalidation Samples, Continued.

	Revised WRNA: Revalidation Sample				
	Arrests	**Conv.**	**Ret.Pris[a]**	**Any Fail**	**Any Fail**
Scale	**Y/N**	**Y/N**	**Y/N**	**Y/N**	**Y/N**
Criminal history		.11*	.20***	.24***	.29***
Attitudes	.09*				
Educational needs[b]					
Educational assets (strength)	.15***				
Employment/financial	.10**	.10**		.13**	.13***
Housing safety	.17***	.11**		.12**	
Antisocial friends					
Anger	.18***	.25***		.21***	.12**
Mental health history					
Depression (symptoms)		.08*		.08*	
Depression (col.)	.08*	.15***		.09**	
Psychosis (symptoms)					
Child abuse					
Adult abuse					
Sex abuse (adult or child)					
Physical abuse (adult or child)					
Substance abuse history	.11*	.12**		.10*	
Substance abuse (current)	.24***	.23***	.12**	.22***	
PTSD[a]					
Parental difficulties					
Parental involvement[b]					
Relation. satisfaction (strength)					
Family conflict	.12**	.12**	.17***	.19***	.09*
Family support (strength)					
Family support (strength)(col.)		-.08*			
Relationship dysfunction		.08*		.08*	
Parental stress (all)		.08*		.14**	.19**
Parental stress (col).			.17***	.13***	.13***
Child abuse (sur.)	.14***			.09*	
Child abuse (col)					
Adult abuse (sur.)		.08*			
Adult Abuse (col)		.08*			
Self-efficacy (strength)		-.10**			
Self-efficacy (col.)		-.11**	-.14**	-.10**	

[a] Rhode Island cases are omitted from return to prison analysis. Data were not available.
[b] Shaded areas indicate that the measure was not tested.
***p≤.01
**p≤.05
*p≤.10

ATTITUDES: The attitudes or criminal thinking scale was related to offense-related outcomes in an inconsistent manner. Yet, it demonstrated an acceptable alpha =.74 reliability rating. Item analysis revealed that one item, offender denies having committed the offense was not contributing to scale validity. Removal of the item, however, did not improve the predictive validity of the scale in the construction validation samples but did in the revalidation sample at the 6 month follow-up. The new scale was related to offense-related outcomes in Kentucky but only for the 6 month follow-up analysis. Alpha for the new scale was .71. Despite modest indication of predictive validity, inclusion of this scale in the cumulative scale forming the stand-alone WRNA detracted from its predictive validity. With respect to cognitive patterns, anger and self-efficacy offered far better contributions to the predictive validity of the total scale. The attitudes scale is included in the needs section of the assessment (Part IV).

EDUCATIONAL SCALES: Neither the educational needs scale nor the educational strengths scale was related of offender outcomes for the sample as a whole. These were standard educational items, and no test items were introduced to improve the scales. The study did, however, provide an opportunity to test whether the scales and scale items were improved by verifying responses through corroborating records and tests. Very few interviewers verified the educational items and perhaps they were not able to do so. Only 116 of the 403 interviews produced verified educational items. It is telling, however, that the educational strengths scales became predictive of outcomes among those cases. Correlations ranged from *r= -.07. not sig.* to *r= -16, p≤.05* at six months. They ranged from -.09 to -.16 at the 12 month point. In this regard, the high school graduation, or lack thereof, was not predictive, however, the receipt of job-related licenses and certificates were especially potent. A new Educational Strengths scale was created which increased the weight of this item, but it failed to improve validity of the item in the construction validation sample until the 12 month follow-up. At that point, the improvement was only noted for the construction validation sample. The revised assessment clarifies the G.E.D and high school items to reduce the possibility of confusion, gives more weight to vocational training, and more strongly recommends corroboration of items. These scales are not included in the cumulative final risk needs scales, but instead appear in part IV.

EMPLOYMENT/FINANCIAL: The employment/financial scale was a robust predictor of outcomes in most community samples. Even so, a number of items did not contribute to the validity of the scale, such as did you own or lease and automobile; did you have a checking account, did you have a savings account. Interviewers also observed that the original items were more appropriate to middle class samples than to samples of justice-involved women. As a result, we tested a number of items that were suggested by practitioners and administrators in the study sites. Those making substantial contributions were:

1. Do you worry about whether you will be able to make ends meet once you are released?
2. Do you live in public housing?
3. Prior to coming here did you have any recent problems like eviction, etc.?
4. Do you live in a household where at least one family member has full-time, year round employment?

As with other scales, a number of interviewers failed to ask the case management (test) questions. The missing data figure, again, was excessive (36.0 percent), so it was necessary to replace these values at the median. While correlations shown in Tables 11 and 12, do not appear to be

appreciably higher than those for the original sample, they were considerably higher among valid cases, with significant findings ranging from *r=.19, p≤.01* to *r=.24 p≤.01*, in the construction validation sample and from *r=.12, p≤.10* to *r=.19 p≤.01*, in the revalidation sample. Results were significant for all sites except Rhode Island. Alpha for the new scale was low .42, reflecting perhaps this missing data problems and some diversity in the items tapped. The employment/financial scale was included in the final risk/needs cumulative scale.

HOUSING SAFETY: As noted in Tables 10 and 11, the predictive validity of the original scale was modest for the total sample. Attempts to improve the scale involved practitioners, who suggested a number of test items for the current study. Two in particular strengthened the scale considerably:

1. During the 18 months prior to your arrest, how many times did you move.
2. Did the police come into your neighborhood a lot.

Unfortunately, both questions were not asked in 43.4 percent of the interviews. Among valid cases, however, the item was highly predictive and contributed to a predictive scale with sufficient internal consistency (alpha=.68). Changes also involved removing the substance abuse item (was your home environment free of substance abuse) from the revised scale. This scale was used in the final risk/needs scale. However, the decision to replace missing values for such a high proportion of cases will be viewed as controversial. Alpha for the new scale was .64, and even with the replacement of missing data, the scale was predictive in all sites.

ANTISOCIAL FRIENDS: The Antisocial Friends scale was strongly predictive of offense-related outcomes in many of the samples tested. In the case of this combined sample, this was especially true at the 12 month point. The only additional item tested, inquired whether any of the woman's close friends on the outside had been on probation or parole? It did not improve the predictive validity. Therefore, this scale remained unchanged, and was included in the final risk/needs scale. Alpha was equal to .74.

ANGER: Across samples, the WRNA Anger Scale tended to be a more potent predictor of offense-related outcomes than the other cognitive variable, e.g., attitudes (antisocial thinking). This could be attributable to the fact that the antisocial thinking scale was more subjective while the anger scale was behavioral. However, earlier focus groups with women offenders informed us that anger ("rage) was a significant factor in their offending. Although the scale had already proven valid, the present study added four test variables to assist an attempt to improve the scale. Three of these were predictive:

1. Have any of these experiences occurred within the past 6 months (excluding self-defense)?
2. Within the past 6 months have you had any times when you think you got too aggressive when something made you angry?
3. Were you angry or upset when you committed the present offense?

These items did not incur the problems with missing data that were so apparent above (missing = 3.2 percent). Addition of the three items to the scale added somewhat to its predictive validity, and alpha for the entire scale was high .80. Results were predictive in all sites except Missouri.

HISTORY OF MENTAL ILLNESS: The History of Mental Illness scale was not predictive of post-prison outcomes. In other writings, we have speculated that this was attributable to the fact that different illnesses were combined into a single scale, when in fact some mental diagnosis were predictive and others were not (Van Voorhis et al., 2010). However, in this case, none of the single items were significantly related to outcomes either. It is possible that a change in status, such as that which might accompany successful treatment, might reduce the predictive merits of this scale. The scale, however, is retained for case management purposes in a separate section of the assessment (Part IV). Alpha for this scale was .78.

CURRENT SYMPTOMS OF MENTAL ILLNESS: In most samples, current symptoms of mental illness tended to be more predictive than the static History of Mental Illness Scale. The prerelease sample, as a whole, found the two current symptoms variables to be predictive in the 12 month follow-up period but not the 6 month period. It is not unusual for risk factors to become more predictive as the follow-up period increases.

> **DEPRESSION:** The original WRNA Depression Scale was weakly correlated with outcomes during the 6 month follow up period, but its predictive validity increased during the 12 month time frame. Two new items were tested, but they did not make any improvements. It was noteworthy, that one existing item, experiencing a loss of appetite, appeared to be detracting from the predictive merits of the scale. When this item was removed, results improved slightly in the construction sample, but not the revalidation sample. The scale was collapsed into low (0), medium (1-4) and high (5) for use in the final risk scale and that too, improved its predictive validity. Results were significantly related to outcome in the three stand-alone sites as well as the two trailer sites at the 12 month follow-up point. Alpha for the new scale was .73.
>
> **PSYCHOTIC SYMPTOMS**: This scale was comprised to two items, which did not predict well for the combined sample during the 6 month follow-up period but became more predictive during the 12 month period in both the construction and revalidation samples. There were no test variables to make any amendments to the scale. The inter item correlation (alpha was inappropriate) was $r=.33, p \leq .01$.

ABUSE-INTERVIEW SCALES: The interview furnished four questions that enabled the creation of 4 abuse scales: a) adult victimization; b) child abuse; c) sex abuse (experienced by an adult or a child); d) physical abuse (experienced as an adult or a child). When samples were combined, these items showed somewhat modest predictions with the outcome measures. However, state specific findings were stronger, especially for the Kentucky samples. With only the original questions asked, there was no possibility of modifying the scales. It is important to note that these items show interviewer effects, where results were stronger for some interviewers than others. Such findings will require changes to training protocols. Three of the four variables were used in the final risk/needs assessment (child abuse, adult victimization and sexual abuse).

PTSD: Four interview items were based upon the Veteran's Administration's Post Traumatic Stress Disorder Scale. The cumulated scale performed very strongly at sites where interviewers did not omit the questions. However, in this prerelease study, data were missing for 27.8 percent of the cases across all items. Missing data may also have affected the results shown in Tables 11 and 12. In this study, there was no way to repair this situation, because all items were affected. A decision was made to keep the scale on the assessment for further study, but it was not included in the risk/needs scale. Alpha for the valid items was adequate (.70).

SUBSTANCE ABUSE: Two substance abuse scales were created for the Women's Risk/Needs Assessments--substance abuse history and current/recent substance abuse. Although they both were valid in most samples, the present study examined the addition of two questions, one for each scale.

> **SUBSTANCE ABUSE HISTORY**: The addition of an item capturing the existence of any current or prior convictions for a substance abuse offense improved the predictive validity in the construction validation sample, but results were modest for the revalidation sample. State specific findings were strong for Kentucky but modest for Ohio and Missouri. Alpha for the scale was substantial=.89.
>
> **SUBSTANCE ABUSE CURRENT**: This scale was often strongly associated with offense-related outcomes. Just the same, the study afforded two opportunities to improve the scale. Not surprisingly an original item asking whether the woman was currently using, detracted from the validity of the scale. Additionally, a test item, do you currently have any feelings that you need to use drugs first thing in the morning, strengthened the correlation. This item, however, was among the test items that interviewers omitted. Missing data was noted for 23.3 percent of the cases and required replacement at the median. Correlations for valid cases were higher than those shown in Table 11 and 12. Alpha for the new scale was .71.

FAMILY OF ORIGIN SCALES: Two scales were created, one measuring Family Conflict and another Family Support. As can be seen in Tables 11 and 12, neither scale was correlated with offense related outcomes. There were however, significant findings in specific samples. At one site in particular interviewers reported that many participants had difficulties conceptualizing siblings and parents in contexts of blended families, separated families, parent figures, etc. Thus, the revised interview is restructured to capture a more nuanced definition of family of origin. In addition test items provided a means of making modest improvements to the scales.

> **FAMILY CONFLICT**: Two items were added to this scale to make modest improvements to predictive validity:
>
> 1. Do any family members have a criminal history?
> 2. Do your parents or any siblings tend to be critical of you when they communicate with you?

It was necessary to replace a large proportion of missing values for the second item (44.8 percent). The change showed a single correlation with outcome in the 12 month construction sample, but strangely was very favorable in the validation sample. The fact that the new scale was predictive in Missouri and Rhode Island and not in others, suggests that interviewer confusion over the structure of the questions may have been problematic at some sites but not others. Alpha for this scale was low (.33).

FAMILY SUPPORT: The family support variable, noted to have some favorable findings was also not possible to improve. No test variables were entered into the study. Alpha for this scale was .78, but the predictive validity of the scale was limited for the construction and revalidation samples. Results were stronger for the Kentucky and Missouri samples. The scale is collapsed for inclusion in the final cumulative risk/needs scale (low=0-2)(medium=3-4)(high=5).

PARENTING SCALES: Parenting scales were highly predictive in several of the community samples. Three were tested, one pertaining to parental involvement and two to parental stress. The purpose for testing two versions of parental stress/difficulties was to determine whether it would be possible to omit one of the scales to assist efforts to shorten the interview process. The earlier construction validation study used only the survey parental stress scale, and did so with favorable results. The goal in the present study was to substitute that for an interview scale.

PARENTAL INVOLVEMENT: This interview scale was not included in the cumulative risk/needs scale because it was valid only when mothers responded. In other words, non-parents could not be included as 0 on the scale. A test item indicating whether or not the participant was having difficulty maintaining or obtaining custody was determined to be relevant to parental involvement scale. In this case an answer of "no" was indicative of involvement. When this item was added to the scale, improvements in predictive validity were seen in the construction but not the revalidation sample. However, findings were significant in Missouri, Ohio and Rhode Island. Alpha was .74 for the revised scale.

PARENTAL DIFFICULTIES: The attempt to create an interview scale to substitute for the survey scale was not successful. Even in cases where this scale was predictive, the survey scale (below) was superior. Test items were missing for 36.5 percent of the 342 women with children and the missing values affected all of the items proposed for the scale.

PARENTAL STRESS: Correlations were observed for the entire sample, and no items were observed to be detracting from the scale. There are however, contradictions between the interviewer's indications of whether the woman has children under 18, and the women's indication on the survey. As a result, the scale was keyed to the interviewer's indication of whether the woman had children. Modifications will be made to the training protocol, to recommend that the interviewer determine that the woman has had at a period of ongoing contact with any children who are 18 or younger at the time of

the interview. Correcting for this must be done during the scoring/research process. Therefore, the questions do not pertain to women who have never had a period of ongoing contact with any children who would have been under 18 at the time of the interview. These women and other non-parents are scored as zero on this scale. The scoring steps are as follows:

1. Total the scale items for women who have children under 18 with whom they have had contacts with.
2. For that group of women, replace any missing cases at the median (13). In this sample, 29 women (8.7 percent) of 334 women who were eligible to answer questions on the survey did not do so.
3. Once the first two scoring steps have been completed, non-parents are entered into the scale as 0.

When this scale is added to the total risk/needs scale, it was collapsed into high, medium and low values:

Low (0) = 0-12
Medium (1) = 13-20
High (2) = 21+

The parental stress scale was predictive in all samples except Rhode Island. Alpha was high (.82).

INTIMATE RELATIONSHIP SCALES: For a number of reasons, the original relationship scales, relationship satisfaction, and relationship dysfunction, seldom correlated with post-release outcomes. Moreover, obtaining participants' responses to these questions incurred a number of difficulties. First, interviewers reported that women were very guarded in their discussions of significant others. As a result, the relationship items had more missing values than other items, even when they were not case management or test variables. Second, researchers observed that interviewers sometimes interjected their own evaluations of whether the woman was actually involved in an appropriate relationship. Interviewers would then alter survey results accordingly. Third, a number of items produced findings that some would find counterintuitive. For example, women involved in significant relationships tended to be slightly more likely to recidivate than those not involved in significant relationships; marriage was not a source of resilience for returning women; and long term relationships did not impact recidivism.

Even so, item analysis revealed ways that the scales could be reconstructed. For example, a factor analysis of the survey relationship dysfunction scale found that it was two dimensional, describing satisfaction as well as dysfunction and criminogenic ties. The revised relationship scales, therefore, more clearly delineated relationship satisfaction and dysfunction.

To reduce the possibility of interviewer bias and improve the privacy of the questions, data for both revised scales will be collected in the survey portion of the assessment. The few

remaining interview relationship questions collect necessary information on marital status and whether or not the woman has a significant other, but these items do not contribute to any scales. The scales are as follows:

RELATIONSHIP DYSFUNCTION: The items comprising this scale are as follows:

1. Do you find yourself more likely to get in trouble with the law when you are in a relationship than when you are not in a relationship?
2. Do you tend to get so focused on your partner that you neglect other relationships and responsibilities?
3. Have partners been able to convince you to get involved in criminal behavior?
4. Do you feel okay about yourself when you are not in a relationship. Or if in a relationship: Would you feel okay about yourself if you were not in a relationship?

Alpha for this scale was low (.61) but improves to .73 without the fourth item, Do you feel okay about yourself when you are not in a relationship? This may reflect the fact that the fourth item was collected by interview rather than survey. As a test item, it incurred a large portion of missing data (43.6 percent). An examination of table 11 and 12 show that the new scale shows improved predictive validity within the construction validation sample at 12 month, but less impressive improvements were shown for the 12 month revalidation sample. The scale does, however, improve the overall predictive validity of the scale in Ohio and Rhode Island.

RELATIONSHIP SATISFACTION: The first three items on the original survey spoke to a sense of satisfaction in intimate relationships. These were added to a fourth item from the interview. Alpha was equal to .70.

1. In general would you describe these relationships as supportive and satisfying?
2. Do you get into relationships that are painful for you?
3. Is your current relationship satisfying to you (i.e., does it make you happy at the present time? [If no significant other, indicate NO.]

As shown in tables 11 and 12, the scale was not predictive of outcomes, except at the Kentucky site. Unfortunately, it may not be possible to create a predictive relationship satisfaction scale among women offenders. Among offender populations, intimate relationships and supportive relationships may also imbed antisocial influences. Scales of this nature must disentangle supportive dimensions from antisocial ones. A similar problem occurred with a social support scale we attempted to create during an earlier phase of this research. Support is often coming from antisocial others.

SELF EFFICACY: The self-efficacy was a well-established Rosenberg Self-Efficacy scale (Alpha=.90) that we did not wish to make improvements to. For addition to the cumulative risk/needs scale, the scale was collapsed into high (24+) and low (0-23). values.

CHILD ABUSE SURVEY SCALE: The validity of the child abuse survey scale appeared to vary by sample. It was possible to reduce the size of the scales by three items. The following were removed: a) pushed/shoved you; b) bent your fingers/twisted your arm, and c) burned/scalded you. Even so, correlations were modest and it was difficult to identify a cut-points to collapse the scale in a manner that would have consistent results across samples. Alpha for the new scale was .96.

ADULT ABUSE SURVEY SCALE: The scale was not strongly associated with outcome measures. Attempts to improve the scale involved deleting two items: a) scratched you; and b) Bent your fingers/twisted your arm. This improved results for only the construction validation sample and for only the Rhode Island sample. As with the child abuse scale finding uniform cut of points was difficult.

It is possible that the modest results for both the child abuse and adult abuse survey scales could have been attributable to their location at the end of the interview and perhaps to some degree of participant fatigue in the face of intense subject matter. Alpha was high (.96).

Constructing a Final Stand-Alone Assessment and Trailer (WRNA-T)

The final stand-alone WRNA scale consisted of the following individual risk/need domains:

Criminal history
Employment/financial
Housing safety
Antisocial friends
Anger
Depression (collapsed)
Psychosis
Child abuse (interview scales)
Adult abuse (interview scales)
Sexual abuse (interview scales)
Current substance abuse
Relationship dysfunction
Parental stress (collapsed)

Additionally, the following strengths were subtracted from the total scale:

Family support (collapsed)
Self-efficacy (collapsed)

Construction of these revised assessments involved examination of several combinations of variables in a construction validation study designed to maximize the predictive validity of the total scales. Scales which failed to contribute to predictive validity were not included, even if psychometric properties of the scale were adequate. Most of these, however, were included in Part IV, assessment of additional needs. Following tests conducted on 6 and 12 month follow-up measures, the total scale was revalidated in a revalidation sample and in the separate sites.

Table 13 shows results of tests conducted on the construction and revalidation samples. Correlations for the construction validation study, as expected, were high, especially for the collapsed levels, where correlations met or surpassed a standard of r=.27 for 7 of the ten tests. The relationship between the stand alone scale and ANY OFFENSE-RELATED FAILURE, the measure considered to be the most uniform measure of outcome across sites, (Pearson's r) was *.29, p ≤. 01* at 6 months, and r=.32*, p ≤. 01* at 12 months. On revalidation, some shrinkage in predictive validity was seen, however, the 12 month results for ANY OFFENSE-RELATED FAILURE remained strong.

AUC values seldom reach a standard of .70 in either the construction or revalidation tests, except on measures of CONVICTIONS at 12 months. This may be attributable to the truncated follow-up period. Results for the analysis of 12 month outcomes are stronger than those for 6 months. This phenomenon is the result of improvements in outcome base rates over time. In most recidivism research such improvements continue until the 2 or 3 year follow-up point. As such two years is the scientific standard for community classification studies. This was clearly the case in the 2004-2008 construction validation research. The results shown in table 13, are likely to be somewhat attenuated by the truncated follow-up period. However, they are somewhat stronger than results shown at the one year point in the 2004 to 2008 research (see

Van Voorhis et al., 2008; Van Voorhis et al 2010). This present study also incorporates more of the specific risk/need scales into the final cumulative scale.

Additional confidence in the final stand-alone assessment can be seen in Table 14, where the final assessment was tested for specific sites. Results were especially strong at the 12 month point for Kentucky and Missouri (on relevant measures of returns to prison, any offense related failure, and any failure, where AUC values often surpassed .70.

Selection of scales for the WRNA-T followed a similar procedure, except scales deemed to be redundant to similar scales on the LSI-R (e.g., family support) were not included unless they provided a gender-responsive definition of a similar scale, e.g., housing safety and employment/financial. The cumulative scale for the WRNA-T consisted of the following risk/need domains:

Employment/financial
Housing safety
Anger
Depression (collapsed)
Psychosis
Child abuse (interview scales)
Adult abuse (interview scales)
Sexual abuse (interview scales)
Relationship dysfunction
Parental stress (collapsed)
Self-efficacy (collapsed) was subtracted from the total.

Tables 13 through 14 also show results for the WRNA-T. On Table 13, results for the revalidation sample showed very little shrinkage from the findings for the construction validation sample. Table 14 shows data from sites testing the stand-alone assessment. Those sites furnished an opportunity to test the WRNA-T scales, although they could not be added to an alternative

Table 13. Bivariate Relationship between Revised WRNA Assessment and 6 and 12-Month Recidivism, across Construction and Revalidation Samples.

	Arrests		Conv.[a]		Ret Pris[b]		Offense Fail		Any Fail	
Scale	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**
Construction Validation Sample 6 Months										
WRNA Stand-alone Assessment (N=201)	.22***	.66	.33***	.79			.23***	.66	.23***	.64
Levels (N=201)	.28***	.69	.29***	.74	.09*	.60	.29***	.69	.27***	.64
WRNA-T: Trailer (N=312)	.17***	.61	.34***	.74			.21***	.62	.19***	.60
Construction Validation Sample: 12 Months										
WRNA Stand-alone Assessment (N=194)	.24***	.65	.25***	.70	.15**	.64	.29***	.67	.24***	.64
Levels (N=194)	.29***	.67	.24***	.68	.16***	.62	.32***	.67	.27***	.64
WRNA-T: Trailer (N=302)	.19***	.60	.26***	.70			.20***	.61	.18***	.59
Revalidation Sample: 6 months										
WRNA Stand-alone Assessment (N=201)	.18***	.62	.22**	.67	.12**	.59	.22***	.63	.21***	.61
Levels (N=201)	.16***	.60	.15*	.60	.13*	.60	.23***	.62	.24***	.63
WRNA-T: Trailer (N=312)	.20***	.62	.13**	.60			.19***	.61	.16***	.58
Revalidation Sample: 12 months										
WRNA Stand-alone Assessment (N=201)	.20***	.62	.29***	.74	.16***	.61	.26***	.65	.18***	.61
Levels (N=201)	.18***	.61	.26***	.70	.18***	.62	.28***	.65	.22***	.62
WRNA-T: Trailer (N=298)	.18***	.60	.23***	.67			.21***	.61	.14***	.68

[a] Missouri cases are omitted from analysis of 6 month conviction data. No convictions had occurred by that point in time.
[b] Rhode Island cases are not included in any analyses of returns to prison. Data on returns to prison were not available.
***p≤.01
**p≤.05
*p≤.10

Table 14. Bivariate Relationship between Revised WRNA Assessment 6 and 12-Month Recidivism, across Jurisdictions.

	Arrests		Conviction		Return to Prison		Offense-Related Failure		Any Failure	
Scale	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**
6 Month Follow-up for Revised Stand-Alone WRNA										
Missouri Revised WRNA (N=195)					.23***	.72	.19***	.64	.32***	.68
Missouri-Levels N=195)					.24***	.71	.23**	.65	.36**	.69
Kentucky Revised WRNA (N=36)	.38***	.73	.31**	.74	.31**	.70	.41***	.75	.41***	.75
Kentucky-Levels (N=36)	.46***	.78	.37***	.81	.37***	.74	.49***	.79	.49***	.79
Ohio Revised WRNA (N=172)	.25***	.65	.27***	.72	-.10*		.19***	.60	.14**	.58
Ohio-Levels (N=172)	.23***	.63	.22	.68			.17***	.59	.14**	.58
6 Month Follow-up Revised WRNA-T Scales										
Missouri Revised WRNA-T (N=195)					.12**	.62	.10*	.57	.23***	.63
Kentucky Revised WRNA-T (N=36)	.34**	.72	.28**	.73	.28**	.68	.35***	.72	.35***	.72
Ohio Revised WRNA Trailer (N=172)	.25***	.66	.24***	.68	-.12*		.18**	.60	.13**	.57
12 Month Follow-up for Revised Stand-Alone WRNA										
Missouri Revised WRNA (N=187)			.20***	.82	.27***	.70	.23***	.64	.22***	.63
Missouri-Levels (N=187)			.19***	.78	.30***	.68	.25***	.63	.27***	.65
Kentucky Revised WRNA (N=35)	.44***	.74	.48*	.81	.38**	.71	.48**	.76	.43**	.72
Kentucky-Levels (N=35)	.49***	.77	.45***	.78	.41***	.74	.53***	.79	.49***	.76
Ohio Revised WRNA (N=169)	.30***	.67	.24***	.67			.25***	.64	.21**	.62
Ohio-Levels (N=169)	.27***	.64	.19***	.62			.23***	.62	.19***	.59
12 Month Follow-up for Revised WRNA-T Scales										
Missouri Revised WRNA-T (N=187)					.14**	.60			.13**	.58
Kentucky Revised WRNA-T (N=35)	.30**	.67	.42***	.80	.25*	.63	.33**	.68	.29**	.66
Ohio Revised WRNA-T (N=169)	.26***	.64	.22***	.65			.21***	.61	.18**	.59

***p≤.01
**p≤.05
*p≤.10

gender-neutral assessment such as the LSI-R or Northpointe COMPAS. It can be seen, however, that the WRNA-T scale was typically significantly related to the outcome variables. Depending on the outcome measure, correlations for Ohio, Kentucky and Missouri often surpassed .20 at the 12 month point. Results are not compared to a higher standard of correlation (r=.27) or AUC level (.70), because the WRNA-T scale, alone, is not intended to serve as a complete assessment; it is missing the gender-neutral scales which would be expected to add to the predictive validity of the tool.

Table 15, shows results for Ohio and Rhode Island, where WRNA-T variables were added to the LSI-R. Results for the WRNA-T, alone (second row for each site) were somewhat better for Rhode Island than Ohio[11], but in both cases, addition of the WRNA-T items to the LSI-R improved the predictive validity of the LSI-R. Moreover, partial correlations show that for all tests conducted the improvement attributable to the WRNA-T was statistically significant.

Table 15 also shows a number of instances where correlations for the WRNA-T plus the LSI-R (third row) are lower than those for the trailer alone. This was not intended, but likely occurred because associations between the LSI-R and the outcome variables were not sufficiently strong. Although this was unanticipated, it implicates the quality of the LSI-R at the sites where the LSI-R data were collected. We do not fault the LSI-R in this regard, as it has sustained extensive validation among women offenders (e.g., see Smith, Cullen & Latessa, 2009). However, WRNA interviewers were trained immediately prior to the administration of the WRNA assessment. In contrast, the LSI-R data were obtained from existing agency records, where staff training may have lapsed, or the time between the assessment and the collection of the data may have been long enough to render assessment results somewhat outdated.

[11] This may be attributable to the fact that the Ohio sample evidenced very few low risk offenders. In all likelihood, they had been effectively screened out from admission to the CBCF, and the addition of the WRNA scales did little to change the fact that very few women were classified as low risk in this sample.

It can also be seen that correlations between collapsed levels and outcome measures were not as high as those for the total LSI and WRNA-T. This reflects some difficulty in setting uniform level cut-points for the two sites. That is, they could have been optimized for each site, but they were not. As a result, the score sheet for the revised WRNA-T recommends that sites re-examine the cut-points and optimize them for their own jurisdictions.

Changes to the WRNA Assessment

The present study and similar research in probation and institutional settings, have resulted in a streamlined assessment that reduces the number of interview questions from 145 to 116. Survey questions have been reduced from 69 to 36. Specifically, the following improvements have been made:

1. As per one of the goals of the current study, the CRIMINAL HISTORY scale is much more predictive of outcomes. It was enhanced by the addition of two new items, and the use of official prior offense measures rather than self-report.

2. The EMPLOYMENT/FINANCIAL scale is revised to tap measures of poverty rather than middle class referents (e.g., do you own an automobile).

3. Enhancements were made to the following additional scales which affected improvements in predictive validity: a) housing safety; b) anger, c) current substance abuse, d) parental involvement; e) depression; f) substance abuse history; g) current substance abuse; and h) relationship dysfunction.

4. The number of case management questions has been reduced, because many were in place to serve as test items. When the study found that they did not improve specific risk/needs domain measures, the questions were omitted, unless they were considered to be important sources of additional information. For example, the question, “Are you experiencing any suicidal thoughts”, is necessary for risk management whether it contributes to a scale or not.

5. Abuse measures are obtained from the interview and not the survey. Survey measures are omitted for several reasons. First, it was not possible to find cut-points for the child abuse and adult abuse scales that were valid across sites; Second, some interviewers noted that the survey questions, which listed specific forms of abuse, were emotionally more difficult for some of the participants, because they were more explicit than the interview questions; Third, the survey questions added considerably to the time needed for the assessment; their omission would contribute to the efficiency of the interview; and Fourth the interview questions tap sexual abuse, which appeared to be highly predictive. The omission of the survey questions does, however, place more responsibility on interviewers to develop rapport and trustworthiness as interviewers. State reports noted a number of instances of interviewer effects, where the abuse items obtained from some interviewers were not predictive.

Most importantly, the current study shows that the revised WRNA and the WRNA-T works across settings. Earlier versions were "fit" to sites, because it was not possible to develop an assessment on a large sample of participants at the outset. Data were not available at a single point in time, so that a large construction and revalidation sample could not be conducted first and then applied to specific sites as in the present study. This study has greatly improved the uniformity of the tool. Sites will vary somewhat in terms of the specific risk/need domains that are predictive, and that is typical of dynamic risk/need assessments, but the total scale is predictive in all sites and cut-points for larger scales including the total risk scale are uniform.

The assessment protocol will, however, encourage sites to develop their own cut points for differentiating low, medium, and high levels of risk. Doing so, without changing the specific scales and the composition of the total scale, often improves the validity of risk assessment instruments, regardless of which assessment is used.

Table 15. Bivariate Relationship between LSI-R and Revised WRNA-T and 6-Month Recidivism, across Jurisdictions.

	Arrests		Conviction		Return to Prison		Offense-Related Failure		Any Failure	
Scale	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**	**Y/N**	**AUC**
6 Month Follow-up for LSI-R and WRNA-T										
Ohio LSI-R (N=137)							.13*	.58	.11*	.57
Ohio WRNA T N=137)	.21***	.61	.23***	.63			.19**	.59	.15**	.57
Ohio LSI-R + WRNA-T (N=137)	.18**	.61	.16**	.61			.19**	.61	.16**	.59
Partial Correlation	.18**		.23***				.15**		.13*	
LSI-WRNA-T Levels	.12*	.56					.14*	.57	.13*	.57
Rhode Island LSI-R (N=222)	.11*	.57	.10*	.58			.10*	.56	.15***	.59
Rhode Island WRNA-T N=222)	.19***	.58	.20***	.66			.24***	.60	.19***	.58
Rhode Island LSI-R +WRNA T (N=222)	.16**	.57	.16***	.63			.18***	.59	.20***	.59
Partial Correlation	.16***		.18***				.22***		.12*	
LSI-WRNA-T Levels	.16***	.59	.13***	.60			.18***	.60	.18***	.59
12 Month Follow-up for the LSI-r and WRNA-T										
Ohio LSI-R (N=134)	.16**	.60		.54	.14**	.66	.23***	.65	.19**	.62
Ohio WRNA-T N=134)	.17**	.58	.20***	.61			.19**	.59	.18**	.58
Ohio LSI-R + WRNA-T (N=134)	.20***	.61	.14**	.59			.26***	.64	.22***	.62
Partial Correlation	.12*		.20**				.12*		.12*	
LSI-WRNA-T Levels	.16**	.58			.12*	.60	.22***	.61	.20***	.61
Rhode Island LSI-R (N=209)	.14**	.57	.14**	.59			.18***	.60	.21***	.62
Rhode Island WRNA- T N=209)	.22***	.60	.28***	.68			.27***	.63	.20***	.59
Rhode Island LSI-R +WRNA-T (N=209)	.20***	.59	.23***	.65			.25***	.62	.24***	.62
Partial Correlation	.18***		.25***				.21***		.11**	
LSI-WRNA-T Levels	.20***	.60	.19***	.62			.23***	.62	.22***	.61

***p≤.01
**p≤.05
*p≤.10

Conclusion

Not withstanding these contributions, there are some necessary precautions to be taken in understanding these findings. First, with the exception of Rhode Island, where the assessment was used for case planning for all inmates, the study samples are still rather small. Sufficient statistical power for a study where construction and revalidation samples are desirable would typically require about 800 cases. The present study amassed data on 403 cases for gender-neutral variables and 626 cases for gender-responsive variables. This necessitated a boot-strap approach where scales were developed in a construction validation sample and retested in a revalidation sample as well as in state-specific samples.

Second, one of the goals of this study was to refine the assessment in ways that would sharpen its predictive validity. To do so, additional questions were added to the interview and tested during the present analysis. In spite of training interviewers to the necessity of asking these test items, several did not. As a result, the missing data problem on a few of the case management/test items required median replacements for as many as 30 percent of the cases on some items. As a result, the decision attenuated the validity of the final scales. This was shown rather definitively when the cases without missing data were tested. Results for those cases were superior to those for the sample as a whole. It should be noted, however, that the missing data problem affected very few test items.

Third, dynamic risk/need scales formed on the basis of interviews conducted in prison could change considerably during the post-prison supervision time frame. The scales were designed to be dynamic and were highly likely to change in the face of such a dramatic change in the participant's environment. This too, complicated the task of developing a pre-release assessment and may have attenuated the overall predictive validity of the tool. Evidence of this

possibility is seen in a comparison of the Kentucky findings to those for Ohio, Missouri, and Rhode Island. The Kentucky interviews were conducted in home communities after the participants had been released. Interviews for Rhode Island, Missouri, and Ohio were conducted in pre-release units of the prison. Correlations for the Kentucky sample were much higher than those for the other sites.

Fourth, the follow-up time period for the present study is 12 rather than 24 months. The earlier studies found more impressive results at 24 months than at 12, and that is a standard observation for both program evaluations and prediction studies. Limited base rates are known to attenuate findings, and longer follow-up periods improve base rates, which in turn tends to improve predictive validity coefficients.

Fifth, though not shown in these analyses, results varied considerably from interviewer to interviewer. Separate analyses for Rhode Island, Missouri, and Iowa (studied as a probation site) found quite clearly that some interviewers produced data which achieved lower predictive validity coefficients than others, especially on sensitive scales pertaining to abuse, trauma, and relationships. Further examination of these findings showed that these interviewers incurred more missing data and were known by their colleagues to have been conducting their interviews too quickly. There are implications for both training protocols and staff selection.

Finally, in many tests, results for the LSI-R trailer were not as favorable as those for the WRNA stand-alone instrument. Table 15 shows a number of instances where the validity of the WRNA-T, found to evidence acceptable predictive validity on its own, became more limited when added to the LSI-R scores. That is validity was "pulled down" by the LSI-R rather than the other way around. As was explained earlier, however, LSI-R data was extracted from correctional files and, in one sample, may have been somewhat dated. Additionally, interviewers

for the WRNA assessments were trained immediately prior to data collection. In contrast a number of state officials observed that many of the LSI-R interviewers were due to be retrained. Dynamic assessments such as the WRNA and the LSI-R require careful monitoring of quality assurance; validity of either assessment is likely to diminish when quality assurance becomes lax.

Even with these limitations, results are in keeping with those typically seen at a 12 month follow-up. The study has succeed in producing a shorter assessment, which finds on re-examination, that more of the gender-responsive variables are predictive than those observed during the original 2004-2008 research. Most importantly, we have much more confidence in the stability of the assessment, because it now is seen to be predictive across several jurisdictions rather than in the one, earlier, construction validation study which sampled only 149 cases.

References

Andrews, D. A.& Bonta, J. (1995).*Level of Service Inventory-Revised.* North Tonawanda, NY: Multi-Health Systems.

Andrews, D. A.& Bonta, J. (2010).*The psychology of criminal conduct* (3rd Ed.). Cincinnati, OH: Anderson.

Belknap, J. (2007). *The invisible woman: Gender, crime, and justice* (3rd Ed.). Belmont, CA: Thompson Wadsworth.

Blanchette, K. (2004). Revisiting effective classification strategies for women offenders in Canada. *Feminism & Psychology, 14,* 231-236.

Blanchette, K. & Brown, S. L. (2006).*The assessment and treatment of women offenders: An integrative perspective.* Chichester: Wiley.

Bloom, B., Owen, B., & Covington, S. (2003). *Gender-responsive strategies: Research practice and guiding principles for women offenders*. Washington, DC: USDOJ, National Institute of Corrections.

Brennan, T.& Austin, J. (1997).*Women in jail: Classification issues.* Washington, DC: USDOJ, National Institute of Corrections.

Brennan, T. (1998). Institutional classification of females: Problems and some proposals for reform. In R. T. Zaplin (Ed.) *Female offenders: Critical perspectives and effective interventions* (pp. 179-204). Gaithersburg, MD: Aspen.

Brennan, T., Dieterich, W., & Oliver, W. (2006). *COMPAS: Technical manual and psychometric report Version 5.0.* Traverse City, MI: Northpointe Institute.

Chesney-Lind, M. (1997). *The female offender: Girls, women, and crime.* Thousand Oaks, CA: Sage.

Daly, K. (1992). Women's pathways to felony court: Feminist theories of lawbreaking and problems of representation. *Southern California Review of Law and Women's Studies, 2,* 11-52.

Farr, K. A. (2000). Classification for female inmates: Moving forward. *Crime and Delinquency, 46,* 3-17.

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

Morash, M. (2006). *Understanding gender, crime, and justice*. Thousand Oaks, CA: Sage.

Morash, M. (2010). *Women on Probation and Parole: A Feminist Critique of Community programs and Services.* Lebanon, NH: Northeastern University Press.

Morash, M., Bynum, T. S., & Koons, B. A. (1998). *Women offenders: Programming needs and promising approaches.* Washington, DC: USDOJ, National Institute of Justice.

Reisig, M. D., Holtfreter, K., & Morash, M. (2006). Assessing recidivism risk across female pathways to crime. *Justice Quarterly, 23*, 384-405.

Van Voorhis, P. (2012). On behalf of women offenders: Women's place in the science of evidence-based practice. *Criminology and Public Policy*, 11:111-145.

Van Voorhis, P. & Presser, L. (2001).*Classification of women offenders: A national assessment of current practices.* **Washington, DC: USDOJ, National Institute of Corrections.**

Van Voorhis, P., Brushett, R. & Bauman, A., (2012). Revalidation of the Women's Risk/Needs Assessment: Results for the Missouri Department of Corrections, Probation, Prison, and Prerelease Samples. Cincinnati OH: Center for Criminal Justice Research.

Van Voorhis, P., Bauman, A., & Brushett, R. (2012). Revalidation of the Women's Supplemental Risk/Needs Assessment: Results for the Rhode Island Department of Corrections, Prison and Pre-release Samples. Cincinnati OH: Center for Criminal Justice Research.

Van Voorhis, P., Wright, E., Salisbury, E., & Bauman, A. (2010). Women's risk factors and their contributions to existing risk/needs assessment: The current status of a gender-responsive supplement. *Criminal Justice and Behavior*, *37*, 261-288.