

Executive Report
Validation of the Risk/Needs Assessment
for use in New Mexico

Prepared by:
The Center for Applied Research and Analysis,
The Institute for Social Research,
The University of New Mexico

Paul Guerin, Ph.D.
Ray Liedka, Ph.D.
Kristine Denman, M.A.

Prepared for:
Community Corrections Program
Probation and Parole Division
New Mexico Corrections Department

January 2000

Table of Contents

Section One - Introduction	1
Introduction	1
Section Two - Methods	2
Data	2
Dependent variables	2
Independent variables	3
Data analysis	5
Section Three - Initial Assessment Results	6
Introduction	6
Proportion of clients in sample who are re-arrested	6
Subsequent arrests by level of supervision	6
Section Four - Results of Validation Check for the Initial RNA	9
Is the instrument valid?	9
Recommended changes	10
Summary of changes to initial RNA	11
Section Five - Final Reassessment Results	13
Comparison of results	15
Section Six - Discussion	19
Limitations of this study	19
Use of the RNA in the future	19
Steps to be taken next	20

List of Tables

Table 3.1 - Percent of Offenders Arrested after Supervision Began 6

Table 3.2. - Subsequent Arrests by Initial Computed Supervision Status 7

Table 3.3 - Subsequent Arrests by the Computed Risk Portion of the RNA 7

Table 3.4 - Subsequent Arrests by the Computed Needs Portion of the RNA 8

Table 3.5 - Subsequent Arrests by Initial Assigned Supervision Status 8

Table 4.1 - Recommended Changes to RNA Items 10

Table 4.2 - Revised RNA 12

Table 5.1 - Percent of Offenders Arrested After Supervision Ended 13

Table 5.2 - Subsequent Arrest (after supervision ended) by Final Computed Supervision Level
..... 13

Table 5.3 - Subsequent Arrest (after supervision ended) by Final Assigned Supervision Level . 14

Table 5.4 - Subsequent Arrests by Computed Risk Portion of Reassessment 14

Table 5.5 - Subsequent Arrest by Computed Needs Portion of the Reassessment 15

Table 5.6 - Suggested Revisions to Final Reassessment 16

Table 5.7 - Revised Reassessment 17

Section One - Introduction

Introduction

The Institute for Social Research (ISR) has completed several reports during its contracts with the New Mexico Corrections Department (NMCD), Probation and Parole Division (PPD) to validate the Risk/Needs Assessment (RNA) currently used by the NMCD PPD. Several analyses and reports (*Status Report: Validating the New Mexico Risk/Needs Assessment Instrument June 1998* and *Validation of the Risk/Needs Assessment for use in New Mexico: Preliminary Findings March 1999*) have been completed as part of the process of validating the RNA and have focused on two different measures of risk: technical violations and termination status. The current report completes this series of analyses and focuses on measures of subsequent arrests (recidivism). This is the only report which focuses on subsequent criminality rather than how the person performs while on probation or parole. Also, unlike previous reports, this analysis includes a check of both the initial assessment and final reassessment.

This report is divided into six sections. First, is this brief introduction section. Second, we discuss the methods used to complete the current analysis. Third, we present the results of the set of analyses which examine how well the initial assessment predicts risk. Fourth, we compare the results of all of the analyses completed from both the current report and prior reports. Included in this section are our recommendations for improving the initial assessment of the current RNA. Fifth, we present the results of the set of analyses which examine how well the final reassessment predicts subsequent arrests. Sixth, recommendations for improving the final reassessment of the RNA are discussed.

Sections two and three deal with the validations of the initial Risk/Needs Assessment while sections four and five are concerned with the validation of the final Risk/Needs Reassessment.

Section Two - Methods

This section describes the methods used to complete the validation of both the initial RNA and final RNA. The initial assessment is administered soon after the client has been placed into a supervision program. The final reassessment is completed at the time the client ends their probation or parole supervision.

The procedure used here is essentially the same as we have used in previous reports. First, we examine how the instrument predicts overall risk. Second, we examine how each item predicts risk. Third, we conduct a series of multivariate analyses to determine the utility of each portion of the RNA.

Data

The sample used for this analysis includes data provided in automated form by the PPD as well as the hard copy data collected by ISR staff.¹ From the initial 2051 cases, 1068 (52%) cases were selected by the SPSS random sample procedure as a construct sample. The other half of the sample will be used to validate any changes made to the RNA. If changes are made to the RNA, before those changes are implemented, they will need to go through the same procedure that was used to assess the validity of the current instrument using the validation sample. Note that this random sample is not the same as the ones used in prior analyses, although certainly some of the clients are in more than one sample. The purpose for generating a new random sample for each analysis is to reduce the likelihood that the results obtained are sample specific. We want to be sure that we are not basing all of our results on some possible peculiarity of a specific sample.

Dependent variables

This analysis includes four measures of recidivism. The first two are used as dependent variables for testing the initial RNA. They are whether there are any subsequent arrests after supervision began and the number of subsequent arrests. The second two are used as dependent variables for testing the predictive efficacy of the final RNA. These variables are whether there were any arrests after supervision ended and the number of arrests after supervision ended. Note that we distinguished between offenses that occurred after supervision *began* and after supervision *ended*. In order to make inferences about the predictive ability of the reassessment, only offenses that occurred after supervision ended could be included. This is because the reassessment cannot predict what has already happened, only what may happen in the future.

We include both whether someone is subsequently arrested and the number of times they are arrested because there may be a difference between offenders who are arrested once versus offenders who are arrested more than once. Additionally, there are sometimes problems with the accuracy of prediction when the criterion measure is binary (arrested or not) and the base rate gets further from 50% (either a large proportion who re-offend or a large proportion who do not) (Gottfredson, 1987). However, concerns regarding the base rate lessen when the criterion is not

¹More detailed information regarding the data is presented in prior reports.

binary. Thus, both binary (arrested or not) and count variables (number of times arrested) are examined.

There are some limitations with the use of subsequent arrests as a measure of recidivism. Subsequent arrests reflect whether the individual is charged with any new offenses after supervision began. This does not *necessarily* reflect whether the person subsequently engaged in criminal activity. It may be that someone engaged in criminal activity but was not caught. Alternatively, a person may have been arrested, but not be guilty.

One way to overcome the first problem, not capturing all subsequent offenses, is to gather self reported delinquency. However, this is a very resource intensive method of obtaining criminality. Further, even self reported delinquency may not be accurate as some people will under-report and others will over-report the amount of criminal activity they have engaged in. Thus, we had to include only official reports of criminal activity. This is also the only information available and is the most valid indicator of criminal activity.

An alternative method that can be used to account for guilt is to use subsequent convictions. However, we discovered that the FBI rap sheets are often missing disposition information. Additionally, subsequent arrests are a more valid indicator of criminal activity and recidivism.

All types of arrests are included (violent, drug, property, traffic offenses, etc.). There has been some concern presented in the literature regarding the inclusion of traffic offenses as a measure of recidivism since traffic offenses are so minor. However, very few (less than 10) of the clients in the construct sample were re-arrested for traffic offenses.

Independent variables

The twenty-three items that are currently used on the initial RNA as well as several other independent and control variables were included as independent variables for assessing the ability of the initial RNA to predict recidivism. These variables are listed below.

- Number of address changes in the last twelve months
- Percentage of time employed in the last twelve months
- Alcohol usage problems
- Other drug usage problems
- Attitude
- Age at first adjudication
- Number of prior periods of probation/parole
- Number of prior felony convictions
- Convictions for property offenses, and
- Convictions for assaultive offense in the last five years.
- Academic/vocational skills,
- Employment,
- Financial management,

- Marital/family relationships,
- Companions,
- Emotional stability,
- Alcohol usage,
- Other drug usage,
- Mental ability,
- Health
- Sexual behavior
- The PPO's impression of the level of the client's needs
- Prior convictions for a violent offense
- Prior convictions for a drug offense
- Whether a weapon was used during the commission of the current offense
- Age at intake
- Whether the offender is living with friends
- Whether the offender is a probationer (versus parolee)
- Gender of offender
- Race of offender
- Length of the follow up period (from beginning date of supervision)

The twenty-four independent variables used for testing the final RNA are the items currently used on the reassessments. Additionally, the other independent and control variables used for testing the initial RNA were also included. Below is a list of the items currently on the reassessment.

- Address changes
- Age at first conviction
- Number of prior probation/parole revocations
- Number of prior felony convictions
- Prior convictions for property offenses
- Percentage of time employed
- Alcohol usage problems
- Other drug usage problems
- Problems with current living situation
- Social identification
- Response to court imposed conditions
- Use of community resources
- Academic/vocational skills
- Employment
- Financial Management
- Marital/family relationships
- Companions
- Emotional stability
- Alcohol usage

- Other drug usage
- Mental health
- Physical Health
- Sexual behavior
- PPOs impression of needs

Data analysis

Several statistical techniques are used to analyze the data to determine whether the instrument performs as expected. The same techniques are used to evaluate both the initial assessment and final reassessment. First, using contingency tables, we examine the outcomes associated with each level of supervision. Outcomes should be worse (a higher proportion of subsequent arrests or mean number of arrests) for clients classified into more intense levels of supervision and better outcomes for clients in lower levels of supervision.

Second, Logistic regression is used to determine whether the instrument predicts whether there are any subsequent arrests. The effectiveness of the risk portion of the instrument, the needs portion of the instrument and other items not currently on the instrument are examined separately.

Third, Poisson regression is used when the dependent variable is a "count" variable, or a measure of the number of times an event occurs. Therefore, this technique is used to assess the effectiveness of each part of the instrument in predicting the number of subsequent arrests.

Finally, a negative binomial regression is used to compare against the results of the Poisson regression. The legitimacy of the Poisson regression model relies on the mean and variance of the conditional dependent variable being the same. **Typically, when there are many zeros-in the current case this means no re-arrest-this property often does not hold.** Thus, we estimate the negative binomial as a check on the poisson analysis.²

²See "Validation of the Risk/Needs Assessment for use in New Mexico: Preliminary Findings," March 1999, for more information regarding the differences between Poisson regression and negative binomial regression.

Section Three - Initial Assessment Results

Proportion of clients in sample who are re-arrested

Table 3.1 illustrates the percentage of the clients who are arrested after supervision began. Approximately 59% of the clients were arrested after supervision began and 46% were arrested after supervision ended. These numbers indicate that the base rate of recidivism for this sample is close to 50-50 for those arrested after supervision ended and is a little further apart for those arrested after supervision began. Most probably it is not far enough to cause substantial problems for accurate prediction.

Table 3.1 - Percent of Offenders Arrested After Supervision Began

	No	Yes
Arrested after supervision began?	41.2%	58.8%

Next, we look at how well the RNA predicts subsequent arrests by the initial supervision status. We examine both the computed and assigned levels of supervision because the classification based on the RNA score can be overridden.

Subsequent arrests by level of supervision

Table 3.2 illustrates the percentage of cases which have subsequent arrests after supervision began by the initial computed supervision status. As expected, the proportion of clients who have at least one subsequent arrest increased as supervision status increased. In other words, clients in higher supervision levels are more likely to be arrested after supervision begins. We also examine the average number of arrests and found the mean number of arrests increases with increasing levels of supervision.

Table 3.2 - Subsequent Arrests by Initial Computed Supervision Status

	Minimum supervision	Medium supervision	Maximum supervision	Total N
No subsequent arrests	64.6%	51.3%	32.8%	379
Subsequent arrests	35.4%	48.7%	67.2%	541
Total N	65	304	551	920
Mean number of subsequent arrests***	.68 (1.31)	1.18 (2.04)	2.01 (2.54)	1.64 (2.36)

p<.001

We then examine the risk and needs sections separately to determine whether one predicts risk better overall than the other. Tables 3.3 and 3.4 provide the results. While both predict recidivism, there are some differences. The needs section appears to more sharply discriminate between those subsequently arrested and those not subsequently arrested than does the risk portion. This is readily seen by comparing the extreme levels of supervision, those assigned to minimum supervision and those assigned to maximum supervision. For the risk portion of the RNA, 68% and 43% of those assigned to maximum and minimum supervision, respectively, were subsequently arrested. But for the needs portion, the figures are 72% and 38% subsequently re-arrested. When the mean number of arrests are compared, a slightly higher average number of re-arrests is indicated for those computed to minimum supervision by the needs portion as compared to the risk portion. However, there is a higher average number of reoffenses for those computed to maximum supervision by the needs portion compared to the risk portion.

Table 3.3 - Subsequent Arrests by the Computed Risk Portion of the RNA

	Minimum supervision	Medium supervision	Maximum supervision	Total N
No subsequent arrests	57.1%	44.0%	32.1%	379
Subsequent arrests	42.9%	56.0%	67.9%	541
Total N	219	243	458	920
Mean number of subsequent arrests***	.88 (1.40)	1.44 (2.26)	2.11 (2.65)	1.64 (2.36)

Table 3.4 - Subsequent Arrests by the Computed Needs Portion of the RNA

	Minimum supervision	Medium supervision	Maximum supervision	Total N
No subsequent arrests	61.7%	45.1%	28.4%	379
Subsequent arrests	38.3%	54.9%	71.6%	541
Total N	115	477	328	920
Mean number of subsequent arrests***	.94 (1.77)	1.38 (2.19)	2.25 (2.63)	1.64 (2.34)

Finally, whether there are subsequent arrests and the average number of arrests by the assigned level of supervision is shown in Table 3.5. Note this is slightly less accurate for clients assigned to minimum supervision as approximately half have at least one subsequent arrest (35% of clients computed as minimum supervision have one or more arrest). A smaller percentage of

medium supervision clients have one or more arrest and nearly the same proportion of maximum supervision clients are recidivists.

Table 3.5 - Subsequent Arrests by Initial Assigned Supervision Status

	Minimum supervision	Medium supervision	Maximum supervision	Total N
No subsequent arrests	50.8%	53.7%	31.4%	379
Subsequent arrests	49.2%	46.3%	68.6%	541
Total N	59	352	509	920
Mean number of subsequent arrests***	1.64 (3.56)	1.06 (1.67)	2.04 (2.51)	1.64 (2.36)

p<.001

Section Four - Results of Validation Check for the Initial RNA

The purpose of this section is to present our conclusions regarding the validity of the initial RNA. The results of all of the analyses which have been completed are compared, and decisions regarding the overall validity of the instrument as well as the validity of each of the items are discussed. Additionally, a draft of a revised RNA to be used at the initial assessment is included.

Is the instrument valid?

One of the first questions we asked when we completed each set of analyses with each of the outcome measures (technical violations, successful completion of probation/parole, whether there were subsequent arrests and number of subsequent arrests) was whether overall the instrument placed offenders into a supervision level that was consistent with the outcome. In general, we discovered that indeed, the instrument did perform as expected. Those classified into higher levels of supervision had worse outcomes while those classified into lower levels of supervision were associated with better outcomes. However, as expected, the instrument did not predict perfectly. For example, 35% of those classified as minimum security were arrested after supervision began. This indicates the percentage of false negatives. Conversely, of the 551 offenders classified as requiring maximum supervision, 182 were not arrested for any other crimes during the follow up period. This may indicate that some of these individuals were supervised at a level higher than was necessary as measured by recidivism (false positives). One question is whether this amount of error is normal for a valid instrument. According to Champion (1994), unofficial error margins of 30% or less are acceptable (p. 206). Is this margin of error acceptable in New Mexico, and if not, can the RNA be improved to reduce this error. In addition, can a revised instrument improve upon this margin or error.

We have also observed that this instrument tends to place the majority of offenders in maximum supervision. It may be that this is appropriate for the State of New Mexico and that most of the offenders who are convicted in this state are high risk. Alternatively, it may be that the instrument is overpredicting the number of offenders who are high risk. The problem with this is high risk offenders use up the state's limited resources. Thus, if the instrument is overpredicting the number of high risk offenders, the state may be using money and manpower on offenders who do not need to be supervised at such a high level. Additionally, the risk and needs portion predict differently. Most offenders are placed into maximum supervision by risk scores, but medium supervision by needs scores.

When we examined each item to determine how well they predicted risk, we found that some items never predicted risk, no matter which outcome measure was used or which statistical technique was used. Clearly, these items are not valid. Other items sometimes predicted risk, depending on outcome or statistical technique. Still others always or almost always predicted risk. These items are clearly valid. Based on these comparisons, the next section addresses the changes we recommend for each of the items.

Procedure used to compare the results

The following procedure is used to determine which items consistently predict risk well and among those that do not, which items need to be changed or possibly eliminated from the RNA. We began by constructing a table which compares the results of each of the analyses. This is attached in Appendix A. This allowed us to compare each of the analyses with each of the criterion measures. By looking at this table, we first identified items which always predict risk. We also looked for categories within these items which predict risk consistently. Second, we looked for items which had specific categories which predict risk. Third, we searched for items which never predict risk. Fourth, we looked for items that never predict risk in the bivariate situation. Since this instrument is additive, if an item does not predict risk in the bivariate, it does not make sense to include it, unless the issues for keeping the item are theoretical, substantive or even political. Fifth, we explored items which had inconsistent results. This included items which only predict some types of risk. Additionally, items which predict all types of risk at least in the bivariate, but are only sometimes significant in the multivariate models are included here as well.

Results of comparisons and recommended changes

The results of this synthesis and recommendations are discussed next. We have included a table summarizing the recommendations for each item (see Table 4.1).

Items which always predict risk

There are no items which predict every outcome measure with every model. However, there are two items which always predict risk as measured by any of the criteria, although not for every model. First, is the number of *prior revocations*. We recommend that this item be included in subsequent drafts of the RNA as it is currently written. Second, the number of *prior periods of probation/parole* is always statistically significant. We recommend that this item not be changed.

Items with specific categories which predict risk consistently

The item measuring *prior property offenses* had one category which predicted each measure of risk consistently. Specifically, prior adjudications for burglary, theft, auto theft or robbery predict risk consistently. However, prior adjudications for worthless checks or forgery does not predict risk. The category which includes both sometimes predicts risk. It is likely that it does because that category also captures the burglary, theft, etc. offenses. Thus, we suggest that this item be altered to include only prior convictions for burglary, theft, auto theft or robbery.

Items which never predict risk

There are two items which never predict risk due to lack of statistical significance or because the direction of the relationship is opposite of that expected. These items are conviction for *assault within the last five years* and *physical health*. Recall that the *assault* item is included on the Wisconsin RNA because it is Wisconsin's policy to place offenders with a recent assault conviction into maximum supervision initially. However, this is not New Mexico's policy. Thus, it makes sense that this item does not predict risk. We suggest that this item be eliminated from the RNA.

As noted earlier, it is expected that, if valid, the scoring of the items will have a linear relationship with the outcome (recidivism). Because the item measuring *physical health* does not, we believe that it should be removed from the RNA. Additionally, very few clients have chronic health problems.

Items which never predict risk in the bivariate analyses

Two items never predict risk in the bivariate situation. These are *marital/family relationships* and *mental ability*. We recommend these items be eliminated from the RNA.

Items which did not predict risk consistently

Included in this category are those items which did not predict every type of risk. It is not surprising that there are items which do not predict every type of risk in every model since the criterion measures differ and are only somewhat related to one another.

Additionally, there are items which may have predicted each type of risk at least in the bivariate situation, but did not always predict risk in the multivariate. Moreover, sometimes only some of the categories of these items predict risk while all of the categories predict other risk measures. Most items had inconsistent results. We begin by discussing those items which predicted only some types of risk. Next, we examine those items that predict all types of risk, at least in the bivariate, but had inconsistent results otherwise.

Items which predicted only some types of risk

The *number of address changes* predicted each measure of risk except whether there is a subsequent arrest (in either the bivariate or multivariate analyses). In some instances, only the category indicating that two or more address changes had occurred is statistically significant. In others, particularly for the models trying to predict the number of subsequent arrests, the key category is no address changes vs. any address change (that is...combine the other two categories). Moreover, bivariate results indicate that one address change does not differentiate between those who pose a risk and those who do not.

Although the amount of *time employed* is always statistically significant in the bivariate, it did not always predict risk in the multivariate models including whether there are any subsequent arrests and the number of subsequent arrests (with the exception of the fourth Poisson model—probably a chance occurrence). Moreover, when it is significant, the item does not always predict as expected (sometimes the direction is opposite than the direction expected). One reason for this may be that the categories should be altered. Another possibility is that the results are confounded by having near duplicate items in the risk and needs portion. We recommend eliminating the duplication and leaving the item as is.

Alcohol usage has inconsistent results. This item never predicts whether the offender successfully completed probation/parole when included in the risk portion of the RNA. However, it did predict risk when included in the needs portion, suggesting this item is not being measured consistently. This item never predicts risk when the criterion is the number of

subsequent offenses. Additionally, when the item does have some statistically significant relationship with the other risk measures, most of the time only the category occasional use rather than frequent use is associated with increased risk when it is in either the risk or needs portion. This suggests that there is not a linear relationship between the scores and risk. Further, alcohol use could be difficult item for officers to accurately determine. Again, this is an item that might be affected by having a near duplicate in the needs portion of the instrument. Based on the results here, we might recommend that this item be eliminated from the RNA. However, we recognize that this item may be substantively important, and the results may be artificial. Thus, we recommend only a single item be kept.

The needs item measuring *employment* does not predict successful completion (except for the category unsatisfactory employment in the first logistic regression model). It predicts technical violations and subsequent arrests in the bivariate analyses only. This item is statistically significant in the Poisson model measuring the number of subsequent arrests, but not the negative binomial model. This item does not appear to predict risk well. This item may also be difficult for officers to measure regarding how the categories are defined and so they may not be used in a consistent manner. We suggest this item be deleted. However, if it is kept, the categories unsatisfactory employment and unemployed should be combined. These categories appear to predict risk similarly and do not need to be kept separate.

The item *financial management* does not prove to be a good predictor of technical violations or number of subsequent offenses. However, the item does predict recidivism in the bivariate and the multivariate analyses. But the key distinction here is between clients with a stable financial history vs. clients with some level of difficulties. The item should be kept, but the final three categories combined.

The item measuring *emotional stability* does not predict whether there are any subsequent arrests or the number of subsequent arrests. While it does not predict technical violations in the bivariate, it does predict it in some of the multivariate Poisson models. This item does not appear to predict risk well. We recommend that this item not be included in the RNA. Like other items emotional stability is a more subjective item and may be difficult to ascertain.

Finally, *sexual behavior* does not predict the number of technical violations, whether there are any subsequent arrests or the number of subsequent arrests. It does predict whether the client completes successfully, but only the category chronic or severe problems is significant and in the opposite direction of that expected. Thus, we propose that this item does not predict risk well and should be dropped.

Items which predicted all types of risk, but not significant in all statistical models

Other drug usage predicts risk inconsistently. While this item predicts every type of risk in at least one of the analyses performed, it is not always significant, nor are all the categories always significant. For example, when the criterion is successful completion of probation/parole, only frequent drug usage predicts risk. However, when significant, this item tends to predict risk as it

is currently written. Therefore, we suggest that this item not be altered, but the duplication across the risk and needs portion be eliminated.

Attitude is often not significant, and when it is, sometimes both categories significantly predict risk, and sometimes only one of the categories predicts risk. However, which category predicts risk is not consistent. It may be that this occurs because relatively few people fall into the last category. It may also be an item that is prone to differential use because of its subjective assessment. We will err on the side of caution, and suggest this item remain the same until the validation check on the revised RNA is completed, at which time we may recommend this item be altered or deleted.

The item *age at first adjudication* is not a significant predictor of risk in every model, but does predict each measure of risk in the bivariate and one of the multivariate analyses. Additionally, in most models, all of the categories are statistically significant. When the outcome is successful completion the less than nineteen years old category is the only significant predictor of risk. However, this item tends to predict risk more often than not as it is currently written. We recommend this item not be altered.

The *number of prior felony convictions* is always significant in the bivariate analyses, but not in the majority of the multivariate analyses. Additionally, only two or more felony convictions are significant when the outcome is the number of subsequent technical violations. Like the previous item, this item may not be the best predictor of risk when other items are included. However, since this item is statistically significant in more than one multivariate analysis, we suggest that this item remain on the RNA unless the validation of the revised RNA indicates that it should be eliminated.

The needs item measuring *academic/vocational skills* often predicts risk but sometimes only low skills was significant. Additionally, there may not be a linear relationship with some of the risk measurements. For example, the average number of technical violations is lower for those with minimal skills as compared to those with low skills. It is likely that the reason this occurs is that there are so few people with low skills. Although there does not seem to be a significant difference between high school and adequate skills, there does appear to be some difference between these two categories. For example, the current analyses indicates that the number of subsequent arrests are higher for those with adequate skills as compared to those with a high school education. Moreover, the coefficient for adequate skills in the logistic model is positive, indicating that those with adequate skills are more likely to have a subsequent arrest as compared to those with a high school education. Since this item does tend to predict each type of risk it should be kept. However, the categories low and minimal skills should be combined.

Next, the item measuring *companions* predicts all types of risk in many of the models. One change that needs to be considered is combining the good support and no adverse relationships categories. One reason for this is that there are so few people that fall into the good support category that these two categories predict risk similarly. Therefore, we suggest these two

categories be combined.

Finally, the item measuring the *PPOs impression of the client's needs* predicts all types of risk, but not all categories predicted risk, nor is it statistically significant in all models. There is no difference between low and minimal needs for most outcome measures. Moreover, those identified as having medium needs are less likely to be re-arrested. This item as it is currently written does not appear to predict risk well. As in other items this may be partially a result of the subjective assessment of the item. It may be more useful to include an **open question** which asks for the PPOs impression of the **type** of risk the client poses rather than the needs the client has. If it is decided that this item should be kept, we suggest that the categories be changed to medium needs or less and maximum needs.

Table 4.1 - Recommended Changes to RNA Items

Item	Recommended changes
Address changes	Combine one address change and two or more changes into a single category
Time employed	No change, but eliminate duplication.
Alcohol usage problems	No change, but eliminate duplication.
Drug usage problems	No change, but eliminate duplication.
Attitude	No change
Age at first adjudication	No change
Number of prior periods of probation/parole	No change
Number of prior revocations	No change
Number of prior felony convictions	No change
Conviction for property offenses	Combine categories to create 2-category version that contrasts those with a prior arrest for burglary, theft, motor vehicle theft or robbery vs. those without.
Conviction for assault in last five years	Eliminate
Academic/vocational skills	Combine low and minimal skills categories
Employment	Eliminate
Financial management	No change
Marital/family relationships	Eliminate
Companions	Combine good support and no adverse relationships categories
Emotional stability	Eliminate
Mental ability	Eliminate
Physical health	Eliminate
Sexual behavior	Eliminate
PPOs impression of client's needs	Eliminate; if kept, convert to open ended question about what type of risk the client poses, or what kinds of needs the client requires.*

*Some items may be substantively important and it may be decided that those items should not be eliminated for those reasons

Summary of changes to initial RNA

First, we recommend that the risk and needs sections of the RNA be combined. In general, the analyses suggested that the needs section is less predictive of risk than the risk portion of the instrument. Moreover, there are particular items which should not be included on the instruments, suggesting a more parsimonious instrument be constructed. The only item not currently included on the RNA which consistently predicted risk is age at intake. Other items which are not able to be measured due to the lack of consistent documentation in the probation/parole files may also be considered for inclusion in subsequent versions of the RNA. For example, previous or current gang involvement may be a good predictor of risk.

The revised RNA questions are presented below. Note that none of the categories are scored. This is because the scoring will have to be determined when the validation of the revised RNA is completed. The validation of the revised RNA will be completed once recommended changes are approved. Additionally, new cutoff scores will have to be determined based on the validation of the revised RNA.

Table 4.2 - Revised RNA

Revised RNA questions	
Number of address changes in the last 6 months	None One or more
Percentage of time employed in the last 12 months	60% or more 40% to 59% Under 40%
Drug usage problems	No interference with functioning Occasional abuse: some disruption of functioning Frequent abuse: serious disruption, needs treatment
Attitude	Motivated to change, receptive to assistance Dependent or unwilling to accept responsibility Rationalizes behavior, negative, not motivated to change
Age at first adjudication	24 or older 20 to 23 years old 19 or younger
Number of prior probation/parole revocations	None One or more
Number of prior felony convictions	None One Two or more

Revised RNA questions

Prior convictions for burglary, theft, auto theft or robbery

None
One or more

Academic/vocational skills

High school or above skill level
Adequate skills: able to handle everyday requirements
Low or minimal skill level

Companions

No adverse relationships
Associations with occasional negative results
Associations almost completely negative

Section Five - Final Reassessment Results

This section presents the results of the validation of the final reassessment of the RNA. Since the automated data received from probation and parole includes only the final reassessment, this reassessment is used for validation. Only one criterion of risk, recidivism, is used to validate the reassessment. Although it may be useful to analyze how well the reassessment predicts other measures of risk like technical violations, because we use the final reassessment this is not possible. One can only predict something that happens in the future so attempting to predict something that has already occurred is not possible and causality cannot be inferred.

Table 5.1 illustrates the percentage of the clients who are arrested after supervision ends. Approximately 46% of the clients are arrested after supervision ends. These numbers indicate that the base rate of recidivism for this sample is close to 50-50 for those arrested after supervision ends.

Table 5.1 - Percent Arrested After Supervision Ended

	No	Yes
Arrested after supervision ended	53.9%	46.1%

Tables 5.2 and 5.3 illustrate the proportion of clients who have at least one arrest after supervision ends by final computed and assigned levels of supervision. The instrument tends to be able to differentiate between clients who will recidivate and those who do not (an increasing proportion of clients recidivate as supervision level increases).

Table 5.2 - Subsequent Arrest (after supervision ended) by Final Computed Supervision Level

	Minimum supervision	Medium supervision	Maximum supervision	Total N
No subsequent arrests	71.7%	58.3%	31.8%	493
Subsequent arrests	28.3%	41.7%	68.2%	422
Total N	311	290	314	915
Mean number of subsequent arrests***	.57 (1.50)	.89 (1.51)	1.96 (2.23)	1.14 (1.88)

$p < .001$

While the percentage of clients who are re-arrested increases with increasing levels of assigned supervision, the differences are not as great as computed supervision levels. For example, the

percentage of clients who are computed to minimum supervision by the RNA and who are subsequently arrested is 28% while 34% of those assigned to minimum supervision are re-arrested.

Table 5.3 - Subsequent Arrest (after supervision ended) by Final Assigned Supervision Level

	Minimum supervision	Medium supervision	Maximum supervision	Total N
No subsequent arrests	66.2%	62.2%	34.2%	494
Subsequent arrests	33.8%	37.8%	65.8%	419
Total N	281	328	304	913
Mean number of subsequent arrests***	.66 (1.41)	.77 (1.42)	1.95 (2.36)	

p < .001

We also assessed how well each portion of the reassessment predicted recidivism overall (see Tables 5.4 and 5.5). We discovered that the two portions predict similarly. For example, both portions are associated with a false positive error rate of approximately 30%. However, the proportion of cases assigned to each level of supervision differs. The risk portion places the greatest proportion (48%) of offenders into minimum supervision, while the needs portion places most offenders in either minimum (40%) or medium (36%) supervision. The risk portion only places 19% of offenders in medium supervision. This difference may account for the difference in the average number of offenses committed by each group. Offenders placed in medium supervision by the risk portion are arrested an average of .96 times while those placed in medium supervision by the needs portion are arrested an average of 1.19 times.

Table 5.4 - Subsequent Arrests by Computed Risk Portion of Reassessment

	Minimum supervision	Medium supervision	Maximum supervision	Total N
No subsequent arrests	69.8%	54.8%	31.0%	497
Subsequent arrests	30.2%	45.2%	69.0%	418
Total N	441	177	297	915
Mean number of subsequent arrests***	.63 (1.52)	.96 (1.52)	1.98 (2.24)	1.13 (1.88)

Table 5.5 - Subsequent Arrest by Computed Needs Portion of the Reassessment

	Minimum supervision	Medium supervision	Maximum supervision	Total N
No subsequent arrests	69.8%	52.3%	30.9%	495
Subsequent arrests	30.2%	47.7%	69.1%	420
Total N	364	331	220	915
Mean number of subsequent arrests***	.60 (1.49)	1.19 (1.84)	1.97 (2.20)	1.14 (1.88)

Comparison of results

The results presented above indicate that there is some error overall- about 30% of offenders are false positives and 30% are false negatives. This suggests that there should be some changes made to the reassessment. Since only one type of risk, re-arrests, is analyzed and so few items are statistically significant in the multivariate models, a slightly different procedure is used to determine whether to change any items. While statistical significance is taken into account, we weighed patterns more heavily than we did when comparing the results of the initial assessment (with the exception of items which are not significant in the bivariate analyses). We are much more conservative with our recommendations since only one type of criterion is measured.

First, we searched for items which never predicted risk, even in the bivariate. We found that three items, *mental ability*, *health*, and *sexual behavior*, are never statistically significant. This indicates that these items do not predict risk well and should be eliminated from the reassessment.

Next, we searched for items which are almost always significant. We found that *age at first conviction* often predicted risk, and is consistent with expectations. That is, as age decreases, the likelihood and number of re-arrests increases. Further, from the bivariate we can see that all the categories perform as expected. We suggest, then, that this item be left in the reassessment, as it is currently written.

Several items are significant in the bivariate and at least one of the multivariate models (usually the Poisson regression). Additionally, these items predicted risk as expected. We suggest that these items be included as currently written. These items are: *current living situation* (although we recommend combining the first two categories to contrast with third), *prior probation/parole revocations*, and *number of prior felony convictions*.

Other items are significant in the bivariate, not in the multivariate, and did not always perform as expected. *Social identification* almost always performed as expected except in Model 3 in the Poisson and Negative Binomial regressions. It is unclear why this occurred. We suggest this item remain on the reassessment as it is currently written. *Employment* for the most part performs as expected. However, unsatisfactory employment and unemployed may need to be combined. First, there are very few people who fall into the latter category. Second, in the Logistic regression, the unemployed category has a negative coefficient, indicating that this category does not predict risk well.

We then examined each item individually. First, we looked at *address changes*. We found that in the bivariate analysis, this item predicted risk as expected. Additionally, although not significant, in the logistic regression, both coefficients for one address change and two or more address changes are statistically significant. This indicates that some number of address changes is associated with increased risk as compared to no address changes.

Bivariate analyses for the item *prior adjudications for property offenses* performs as it did in the

initial assessment. That is, only the item burglary, theft, auto theft or robbery really appears to predict risk. Of some interest is that the category burglary, theft, etc., has a negative coefficient in the logistic regression analysis. This indicates that people with prior burglary, theft, etc., offenses are less likely to re-offend. This result may have been found because the other two categories are not good predictors of risk. We suggest that if this item is kept, only the category burglary, theft, etc., be kept.

Next, the *percentage of time employed* in the last six months is examined. This item performs as expected in the bivariate. This item is not significant in the Poisson models, however. This would suggest that the item should possibly be eliminated, but on the basis of the good pattern of prediction in the bivariate, we merely suggest keeping the item as is, and waiting to see if the validation of the revised RNA will point towards dropping the item.

We then examined the items measuring *alcohol usage*. This item predicts risk in the bivariate. In the logistic regression, the second category, frequent abuse, had a negative relationship with risk. The relationship with risk is opposite the expected direction in the Poisson and negative binomial regressions. We attribute this to the inclusion of a near duplicate item from the needs portion of the instrument. We recommend keeping only one of the two near duplicate items.

Other drug usage predicts risk as expected in the bivariate. It should be noted that there are inconsistencies between the risk and needs portion, however. The average number of offenses differs between all pairs in the needs portion, but not in the risk portion. This item predicts as expected in the first two models of the Poisson and negative binomial regression, but not in the logistic regression. The last category (frequent abuse) has a negative relationship with risk in the logistic regressions. However, since the usual pattern is consistent with expectations, we suggest that this item be left as it is, but that the near duplicate item from the needs portion be eliminated.

The item on *use of community resources* consistently behaved contrary to expectations. In the bivariate analysis, the clients least likely to recidivate are those clients who do not need community resources and those clients who rejected them. The effect of this item in predicting the average number of re-arrests was not significant. Thus, we suggest eliminating this item.

Marital/family relationships performs as expected in the bivariate and is not statistically significant in any of the multivariate regressions. We suggest this item be eliminated.

The item *companions* performed as expected in the bivariate, but often did not perform as expected in the multivariate models. We suggest this item be eliminated.

We then looked at the item measuring *emotional stability*. We found that it predicts risk as expected, but there is no significant difference in the average number of re-arrests in the last three categories. Moreover, the logistic regression models suggest that the last category is not as highly associated with risk as the previous two. The coefficients in all of the models of the Poisson and negative binomial regressions are negative. This item does not appear to predict risk well. We suggest that it be eliminated.

Financial management is significant only in the bivariate, but performs as expected in the multivariate. However, from both the bivariate and logistic regression, it appears that there is little difference between the situational and severe difficulties. Thus, we suggest these categories be combined.

Finally, the *PPOs impression of the client's needs* is statistically significant in the bivariate, and is not significant in the Poisson regression, nor in the negative binomial regression. However, the bivariate analysis suggests that offenders in the first three categories are less likely to be re-arrested and had a similar number of arrests. This suggests that if the item is kept, then these three categories should be combined. But given the lack of significance in the multivariate analysis, we suggest that this item be considered for elimination.

We also suggest that the duplication of the alcohol and drug usage items on the risk portion and the needs portion be eliminated. One reason is that there is some evidence that these items are not coded consistently, leaving us to question whether these items are useful. Second, it makes more sense to only include it on one portion. Further, it may be fruitful to discuss the possibility of combining the risk and needs sections into one measure.

6
Table 5.10 - Suggested Revisions to Final Reassessment

Reassessment item	Recommended changes
Address changes	Consider altering categories to include no address changes vs. one or more address changes
Age at first conviction	No change
Number of prior probation/parole revocations	No change
Number of prior felony convictions	No change
Conviction for property offense	Combine categories to create distinction between those with prior convictions for burglary, theft, auto theft, and robbery vs. those without such prior convictions.
Percentage of time employed	No change (but eliminate duplication)
Alcohol usage problems	No change (but eliminate duplication)
Other drug usage problems	No change (but eliminate duplication)
Problems with current living situation	No change
Social identification	No change
Response to court imposed conditions	No change
Use of community resources	Eliminate
Academic/vocation skills	Combine low skill level and minimal skill level categories

Employment	Eliminate (to avoid duplicating item in risk portion)
Financial management	Combine situational or minor difficulties and severe difficulties categories
Marital/family relationships	Consider elimination.
Companions	Consider elimination.
Emotional stability	Eliminate
Mental health	Eliminate
Physical health	Eliminate
Sexual behavior	Eliminate
PPOs impression of client's needs	Consider elimination.

5.7

Table 5.11 - Revised Reassessment

Revised reassessment risk/needs	
Number of address changes in the last 6 months	None One or more
Age at first conviction	24 or older 20 to 23 19 or younger
Number of probation/parole revocations	None One or more
Number of prior felony convictions	None One Two or more
Percentage of time employed in the last 12 months ..	60% or more 40% to 59% Under 40%
Alcohol usage problems	No interference with functioning Occasional abuse: some disruption of functioning Frequent abuse: serious disruption, needs treatment
Drug usage problems	No interference with functioning Occasional abuse: some disruption of functioning Frequent abuse: serious disruption, needs treatment
Problems with current living situation	Relatively stable relationships Moderate disorganization or stress Major disorganization or stress
Social identification	Mainly with non-criminally oriented people Mainly with delinquent persons
Response to court imposed conditions	No problems of consequence Moderate compliance problems Has been unwilling to comply
Academic/vocational skills	High school or above Adequate skills; able to handle everyday requirements Low or minimal skills causing adjustment problems

Revised reassessment risk/needs

Financial management

Long standing pattern of self sufficiency
No current difficulties
Situational or severe difficulties

Marital/family relationships

Relationships strong/stable
Some disorganization or stress
Major disorganization or stress

PPOs impression of client's needs

Minimum, low or medium
Maximum

Section Six - Discussion

There were several goals to this report. First, we presented the results of the validation of the initial RNA with respect to recidivism as measured by subsequent arrests. Second, the validation of the final reassessment with respect to recidivism was detailed. Third, we suggested revisions to be made to both the initial RNA and the final reassessment. In this final section, we discuss the limitations of the present study and how that impacts our results, we discuss how we envision the way the RNA would be used in the future, and what steps need to be taken next.

Limitations of this study

There are at least two limitations to this study. The first is that the recidivism measure only measures official arrests. Thus, this does not measure all subsequent offenses committed. One problem this may cause is that some offenses may be less easily detected than others. For example, someone who drives under the influence of alcohol and/or drugs may do so numerous times and only be caught once or twice, or never. This particularly effects the reassessment since recidivism is the only measure of risk used to validate this portion of the instrument. This is not an unusual limitation for this type of study and there are few methods to rectify this limitation.

The second problem is the validity of the reassessment is based solely on the final reassessment, rather than any of the reassessments occurring after the initial and before the final. It could be that certain items in the reassessment predict performance while on probation and parole, but not recidivism. Thus some items that we have suggested be eliminated may actually predict other risk measures. Only the final reassessment could be considered because the automated database provided by the Probation and Parole Division only included the last recent reassessment. This is consistent with the protocol outlined by Wisconsin. However, we suggest that all of the reassessments be automated. Not only would it be useful for future validation studies, it would also be of value for offender progress over time for those items that are dynamic.

Use of the RNA in the future

We suggested that the risk/needs assessment be combined into one instrument, eliminating most of the needs items – and in particular the near duplicate items. This validation is based on the statistical prediction of **risk**. Although the needs portion may be very useful for supervision planning, the analyses indicate that statistically, most of these items do not predict risk. The very items that do not predict risk, however, may actually be useful for assessing the level of needs of a client. However, we suspect that rarely will a client require high levels of supervision only due to their needs. Such clients are quite likely to turn out to be risky as well.

Another aspect of some of the needs items recommended for elimination, is that the item categories may not be well defined. For example, the difference between unemployed and unemployable may not always be clear cut. Further, the PPO survey indicates that probation officers do not always feel qualified to assess some of these items. For example, whether the client has sexual behavior problems may be difficult to ascertain. It may be more appropriate to ask whether there is any indication that the client has committed a sex offense. We recognize

that some items may be substantively important, although they are not statistically significant. Particular items that have substantive importance in determining risk should be included on the RNA. We also believe that it would be prudent to include a second independent instrument for measuring needs. We envision this independently developed assessment instrument would be used for supervision planning. Thus, one instrument would be used to assess an offender's risk and place him/her into a supervision level. After the offender's risk is assessed, the needs portion could be used to help determine the offender's appropriateness for certain programs. For example, if an offender poses an extremely high risk, but has low needs as measured by the second instrument, that offender may be placed in ISP. However, if the offender poses high risk and high needs, then the offender might be more appropriate for Community Corrections. Thus, the risk portion would only be used to determine the level of supervision the client should be placed in. The new needs instrument, which could include items currently on the pre-sentence report such as religiosity or socio-economic problems, would be used specifically for supervision planning and placement.

Additionally, other items not currently on the RNA may need to be included. Such items may include gang involvement, whether a weapon is used in the current offense, or any other item deemed important for predicting risk. It should be kept in mind that some items are predictive of risk, but not needs, and that the needs, although important, do not necessarily predict risk. A focus group with PPOs may need to be conducted to specify items that should be included. A comparison of the results from this validation study and the PPO survey may reveal more.

Steps to be taken next

The next step to be taken is a meeting between ISR evaluation staff and PPD staff to discuss the potential revisions to the RNA. Once items to be included have been agreed upon, a validation of these items with the holdout samples needs to be completed. The analyses will indicate which of these new items, old items and old items with revised categories, should remain on the final draft. Once these items are selected, the weights of the categories need to be computed based on the analyses. Third, new cutoff scores to determine the supervision levels will have to be figured. If a needs instrument to be used for supervision planning and placing offenders into special programs is to be constructed, this will have to be completed before implementing the revised RNA. A focus group is recommended to finish this step.

In order to implement the revised RNA, the following steps need to be completed. First, a new training manual will have to be written. Second, a training on how to use the new RNA should be completed. Third, a reliability check will need to be completed by the evaluation staff to ensure that PPOs are all completing the forms as consistently as possible. Finally, the form may be fully implemented. Throughout this process, we may discover that some items need to be refined, thus this will occur as necessary.

Finally, the point at which the instrument is first administered by the PPOs needs to be discussed. We believe that the appropriate point of initial administration is before the client is assigned to a program. This is crucial for the validity of the instrument, and for the planning and

placement form (if this is to be used). It makes no sense to place an offender into a program prior to discovering the risk the offender potentially poses. If an offender is placed into Community Corrections, but measured objectively is really a low risk client, this is a waste of the Division's limited resources. The original Wisconsin protocol mandated that offenders be assessed after being assigned to a PPO. This makes sense when the only program available is either probation or parole. However, since there are now numerous special programs, altering the point at which the instrument is administered is paramount to its utility.

Periodic checks of the validity of the new instrument need to be completed. The instrument could be perfectly valid now, but over time, the instrument could no longer be valid. Essentially, the risk prediction instrument may need to change with the times.