Validation of the Risk/Needs Assessment for use in New Mexico: Preliminary findings

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

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

Prepared for:
Community Corrections

March 1999

Executive Summary

Contracted Scope of Service

- Validate the current Risk/Needs Assessment (RNA) instrument.

Objectives

- Briefly describe the popularity and purpose of statistical risk prediction instruments.
- Describe the creation and implementation of the RNA in Wisconsin and compare this with the implementation of the RNA in New Mexico.
- Determine the percentage of cases in each level of supervision based on the RNA scores. Compare this with the actual assignment of cases into each level of supervision.
- Determine whether the RNA instrument is valid for assessing performance on probation or parole as measured by the number of technical violations the client has documented in their files.

Findings

A. Risk prediction instruments

- Statistical risk prediction instruments are considered an objective, consistent and more accurate method of determining an offender's risk as compared to clinical decision making; research comparing clinical and statistical prediction of risk have verified this hypothesis.
- The purpose of risk prediction devices is to classify offenders into appropriate levels of supervision according to their level of risk.
- Research on the levels of supervision clients receive has found that increased supervision reduces recidivism among offenders who are high risk, but increased supervision among those who are low or medium risk does not impact recidivism. Thus, risk prediction instruments are necessary to help determine where resources should be focused.

B. The implementation of the RNA in New Mexico

- The RNA predicts three levels of supervision while the New Mexico RNA indicates four levels of supervision (minimum, medium, maximum and Intensive Supervision (ISP)).

There is no formal cutoff score for determining the fourth level of supervision, ISP.

- Offenders can be supervised by several other special programs as well, such as Community Corrections, Drug Court and Domestic Violence. There are no formal cutoff scores for these programs.
- The cutoff score for the needs portion of the RNA was changed for use in New Mexico while the risk score cutoffs remained the same.
- Offenders in special case management programs are accepted into the program before the RNA is administered; thus, their level of supervision is determined by program guidelines rather than the RNA score.
- The creators of the RNA purposely weighted the risk item relating to recent assaults to classify an offender at the maximum supervision level if the offender had a recent assault. This was Wisconsin's policy, but may not be New Mexico's policy.

C. Percentage of cases in each level of supervision

- Over half of the clients are classified into maximum supervision, approximately 35% are classified into medium supervision and the remaining are classified into minimum supervision based on the scores from the RNA. Since the distribution of cases is top-heavy, this suggests that the instrument tends to overpredict risk. Alternatively, it could mean that the majority of New Mexico offenders represent a high risk. However, this is less likely.
- The level of supervision that clients are actually assigned to is the same as the level of supervision computed by the RNA in most of the cases (74%) once ISP and maximum supervision are combined. The assigned level of supervision tends to change to the next level up or down (e.g., from maximum to medium) rather than two levels up or down (e.g., from maximum to minimum). Further, more offenders are assigned to levels of supervision lower than that computed by the RNA.

D. The validity of the RNA for predicting performance on probation or parole

- The average number of technical violations is greatest for those who were classified into maximum supervision based on the RNA as compared to those in medium and minimum supervision. The average number of technical violations was slightly higher for offenders classified into medium supervision as compared to those in minimum supervision; however, the difference was not statistically significant.

- Overall, the items on the risk portion of the RNA tend to predict the average number of technical violations better than the needs portion. This is shown both in the bivariate analysis and multivariate analysis.
- There were several items on the RNA that have a relationship that would not be expected if the items are valid. This suggests that either the items are not appropriate, the categories need to be changed, or the weights need to be altered.
- Several items that are not currently used on the RNA were included in one statistical model; it was found that many of these items predict performance on probation or parole suggesting that there are other variables that should be included on the RNA that are not currently used.

Recommendations

- Determine the RNA cutoff score for both ISP and Community Corrections and include it as part of the policy for determining who is eligible for these programs. According to the 1994 edition of the New Mexico Criminal and Traffic Law Manual, ISP is to be used for those clients who "represent an excessively high assessment of risk of violation of probation or parole." Likewise, Community Corrections is to be used for clients who are high risk and very high needs. The RNA should be used to help evaluate whether an offender is eligible for these programs.
- The RNA should be completed prior to admitting offenders into special case management programs, rather than afterwards which is the current procedure. According to policy guidelines, offenders who are being considered for special case management are evaluated for those programs prior to referral or placement into the program. At the time the offender is being assessed for their eligibility based on other criteria (i.e., geographic distance from the program, access to phone for electronic monitoring, and employment) the RNA should be completed as well.
- Although risk prediction instruments are thought to be more objective and accurate than

clinical decision making, the instrument can only work well when it is used the way it was meant to be used. There was some discrepancy found between the coding of alcohol and drug use on the risk portion as compared to the needs portion. While errors will occur, it is of the utmost importance to minimize errors as much as possible, otherwise, the instrument is useless at best, misleading at worst. Thus, whenever the RNA is filled out, it should be double checked for accuracy.

- Related to the previous recommendation, conducting a reliability check to be certain that everyone is filling in the forms appropriately should be considered.
- At least one of the items on the RNA reflects Wisconsin's policy for supervising offenders: a recent assault will automatically score an offender high enough to be placed in maximum supervision. This may not reflect New Mexico's policy and should be altered if it does not.

Future Analyses

While this analysis is important, it is only the first step towards validating the RNA. The next steps that need to be taken are as follows:

- Use a different statistical technique to analyze this data to verify the results presented here.
- Determine whether the RNA predicts the offenders' ultimate termination status (completed or not) from probation/parole.
- Determine whether the RNA predicts recidivism from the subsequent arrest data.

After *all* of these steps are completed, the results of each analysis will be compared. This will tell us whether the instrument is valid. Next, we will need to revise the instrument accordingly and verify the results on the validation samples. Then, the revised instrument will need to be implemented and a subsequent validation check will have to be completed.

Table of Contents

List of Tables and Figures

Chapter One - Introduction

Introduction

The New Mexico Corrections Department (NMCD), Probation and Parole Division (PPD) contracted with the Institute for Social Research to perform a number of research tasks. One of these tasks is to validate the Risk/Needs Assessment (RNA) currently used by the NMCD PPD. The goal of this report is to present the results of the research that has been completed to date with respect to this task.

The validation of the RNA instrument is designed to be a process rather than a single assessment; the current research is the first step in the process. When determining whether an instrument accurately predicts risk, multiple measures of risk should be investigated (Gottfredson, 1987). The Wisconsin RNA was intended to predict not only future offenses but also performance on probation/parole. Thus, there are several possibilities regarding the validity of the current RNA. First, the RNA may predict performance on probation/parole. Second, it may predict whether an offender completes probation/parole. Third, it may predict subsequent criminal involvement. Finally, it may be valid for some combination of these or none of the above.

The current study focuses on one measure of performance on probation/parole: violations of rules. Several possibilities regarding the accuracy of the instrument with respect to rules violations are investigated. First, we consider whether the instrument in its entirety predicts technical violations. Next, we evaluate whether each of the RNA items included predict risk. Third, we examine whether the needs assessment as a whole predicts technical violations. Fourth, we assess whether the risk assessment as a whole predicts technical violations. Finally, we examine whether other variables which are not included on the RNA are better predictors of risk than the RNA items. This research is the beginning of the validation of the RNA and the results are preliminary; conclusions regarding the effectiveness of the RNA should not be based solely on the results presented here.

The report is organized as follows. The first section of this report describes the rise in the popularity of objective risk assessment devices. Next, we provide a background of the Wisconsin Risk/Needs Assessment including its development and its intended use followed by a brief description of the use of the instrument in New Mexico. Next is a discussion about risk prediction in general: what it means and the problems associated with it. Finally, the results of the validation check are presented.

Chapter Two - Background information

This chapter provides a background for the purpose and popularity of risk prediction instruments in general. We then describe the development and implementation of the Wisconsin Risk Needs Assessment. The use of the RNA in New Mexico is then discussed. Finally, we address the meaning of risk prediction and the factors that affect the accuracy of any risk prediction instrument.

The shift towards actuarial prediction

Predictions regarding an offender's risk have been made in the field of corrections since its inception. Judges, probation and parole officers and others make decisions about whether an offender will successfully complete probation or parole, will stay off drugs, will re-offend, and other judgments. There are two methods that have been used to predict risk: clinical decision making (the criminal justice agent determines risk based on their own professional judgements) and statistical risk prediction (the use of some scale to assess risk). Beginning in the 1970s, there was a movement away from the use of clinical decision making towards statistical risk prediction. There are several reasons for the rise in the popularity of statistical risk prediction. First, researchers concluded that a small number of individuals commit a great proportion of crime (Clear, 1988; Jones, 1993). Thus, criminal justice agencies needed an economic way to identify serious, repeat offenders (Jones, 1993). Second, there has been a demand for better, more accurate, and more consistent decision making in the criminal justice system (Clear, 1988; Jones, 1993; NIC, 1980). Studies comparing clinical decisions versus statistical predictions repeatedly found that statistical prediction is much more accurate (Jones, 1993). Third, the use of computers allows agencies to manage information as well as perform statistical analyses in a relatively inexpensive way (ibid). A fourth impetus for the widespread use of risk prediction instruments is that the National Institute of Corrections (NIC) advocated for the nationwide implementation of a workload monitoring system developed in Wisconsin.

Many states adopted the system including a Risk/Needs Assessment (RNA), which is used to classify probationers and parolees into different levels of supervision based on their scores.

Studies examining the effects of different levels of supervision for offenders posing various levels of risk have found that subsequent criminality decreased among high risk offenders who were supervised more closely while the level of supervision has little or no impact on the future criminality of low risk offenders (Jones, 1993; Wagner and Krausman, 1991). Thus, the NIC determined that a risk assessment instrument should be used to help classify offenders, but would not fund the development of new risk assessment instruments due to the prohibitive costs. Rather, they suggested that states adopt an existing instrument, although the instrument adopted did not have to be the Wisconsin RNA. The NIC indicated that whichever instrument was adopted should be validated within a few months of its implementation in a new state to allow revisions to be made as needed. New Mexico is among the states that adopted the RNA (in conjunction with the workload monitoring system) from Wisconsin.

The Wisconsin Risk/Needs Assessment: A Background

In the 1970s, the federal government funded an experimental unit within the Wisconsin Division of Corrections, Bureau of Community Corrections that was intended to lessen the client to officer ratio as well as implement a workload inventory system and specialized caseload for probation and parole officers (NIC, 1980). This resulted in the development of the Risk/Needs Assessment (RNA), as well as the establishment of a management information system, a method to determine the appropriate supervision strategy and a workload accounting of caseloads (Wright, Clear and Dickson, 1984).

Creation of the RNA. Wright et al. describe the development of the RNA as being routine. First, variables that potentially are associated with outcomes were coded from a sample of closed files. Some of these variables were eliminated using a bivariate approach. In other words, those variables that did not have a relationship with the outcome measure were excluded from further analyses. Next, a regression analysis that included all the remaining variables was conducted. Based on the results of this analysis, variable weights were created. For example, the first question on the risk portion of the RNA, number of address changes, carries less weight overall than the third question, alcohol usage. The values for each of the responses for the number of address changes range from 0 to 3; for alcohol usage the values for each of the responses range from 0 to 4. The new weighted model was then validated against another sample of cases. The cutoff levels and supervision classifications were determined from the validation.

A follow up study of the RNA focused on multiple outcomes to determine the predictive ability of the instrument (NIC, 1980). These outcomes include the following: absconsions, rules violations, arrests, misdemeanor convictions, felony convictions and revocations (ibid). They concluded that the assignment of offenders to different levels of supervision based on the RNA decreased the number of subsequent convictions, violations of rules, absconsions and revocations for maximum risk clients and had no detectable adverse affects for low risk clients. Further, they determined that the RNA does predict success or failure on probation/parole (ibid).

Description of the RNA and its implementation. The RNA consists of four parts: the Assessment of Client Risk, Assessment of Client Needs, a coding sheet and a Supervision/Treatment Plan. The Assessment of Client Risk contains 11 items that results in a total risk score which ranges from 0 (no risk) to 52 (high risk). The Assessment of Client Needs contains 12 items, which results in a total needs score which ranges from -7 (indicating strengths) to 60 (high needs). The coding sheet is used to summarize the scores and indicate the assigned level of supervision. Additionally, it is used to collect demographic information, prior felony information, ordered restitution, and county of conviction. The Supervision/Treatment plan allows the probation or parole officer to specify problem areas, behavior objectives, any special conditions, and resources to be used. The level of supervision is determined by the clients total risk or needs score, whichever is highest. However, the level of supervision may be adjusted by the PPO with a supervisor’s approval, if necessary. There are three possible levels of supervision: minimum, medium, and maximum. Each of the supervision levels require a specific number of contacts between the PPO and the client. The minimum level of classification, however, may consist of either a face-to-face visit every 90 days or mail in supervision. The cutoffs for each supervision level in Wisconsin are presented in Appendix A.

The designers of the Wisconsin RNA determined that the initial assessment should be administered within 30 days after admission to probation or parole in order to determine the client’s classification level. The client is reassessed at six-month intervals so that any changes in the client's risk or needs result in a change in the level of supervision, if appropriate. The client is given a final reassessment at termination, which includes the clients' risk and needs scores, as well as the client's termination status. In addition to the initial assessment, the most recent reassessment data is kept on file.

The use of the RNA in New Mexico

New Mexico follows the same assessment-reassessment time line set by the creators of the RNA for offenders that are in "non-special" case management. Clients are supervised at the maximum level until the initial RNA is completed. Clients in the special case management programs are assessed within seven days. However, the RNA does not determine the offender's entrance into the special management program. Rather, whether clients meet the criteria set for the program determines their eligibility; apparently, their RNA score is not one of those criteria. It is unclear what role the RNA plays when the clients are in special management programs.

ISP is one of the special programs which the State of New Mexico uses for clients who are determined to be in need of an exceptional amount of supervision. The cut-off scores for the RNA, however, were not changed to reflect this category despite the fact that the rules of criminal procedure clearly state that ISP is to be used for individuals who "represent an excessively high assessment of risk of violation of probation or parole" (NM 31-21-13.1) and that ISP is a category listed on the RNA instrument.

Community Corrections is another special case management program. According to policy guidelines, offenders who are in this program are high risk and have exceptionally high needs. Again, there are no formal cutoff scores to determine whether clients are eligible for this program.

While reviewing the policy guidelines, we discovered that there were cutoff scores handwritten into the policy that indicated cutoff scores for both ISP and Community Corrections. However, in discussions with Probation and Parole Division Staff, we were informed that indeed, these cutoff scores do not reflect the policy guidelines of the PPD. Thus, there is no systematic method used Statewide to classify offenders into ISP or Community Corrections based on the RNA scores. Additionally, there are other special programs in the State of New Mexico that probation and parole clients may be assigned to, but the RNA does not reflect these either. Drug Court and Domestic Violence programs are other programs that clients may participate in but are not based on RNA scores.

New Mexico also differs from the original Wisconsin protocol in one other way. The cutoff levels for the needs portion of the instrument were changed for use in New Mexico: the cutoff for determining minimum and medium needs is seven points lower. The cutoff score for maximum needs is ten points lower. The cutoff scores for the risk portion of the instrument remain the same (see Appendix A for cutoff scores).

Like the Wisconsin protocol, the database where the assessment-reassessment data is stored follows the same procedures. The initial assessment and most recent reassessment are kept; all other reassessments are deleted.

There are other potential problems associated with the use of the RNA in New Mexico. Although risk prediction instruments are perceived of as objective, value-free tools for predicting risk, there are policy guidelines and value decisions that guide the inclusion of some items rather than others, the use of particular cutoff points, etc. (Jones, 1993). Indeed, at least one item on the RNA instrument reflects the corrections supervision policy of the State of Wisconsin. The final item on the risk assessment is whether there has been a conviction (juvenile or adult) for an assault within the last five years. The points associated with this score are 0 for no and 15 for yes. This scoring was chosen because it was Wisconsin's policy that offenders with an assault would be assigned to the maximum level of supervision, at least initially. Thus, the minimum cutoff score for maximum supervision, 15, is the score for a yes. The State of New Mexico did not change this weighting even though it may not reflect New Mexico's policy. Additionally, no items were added to the RNA instrument.

Prediction

The prediction of an individual's risk is based on the behavior of a group with whom the individual has similarities (Clear, 1988; Jones, 1993). Thus, an individual is classified according to how members of a group tend to act (Clear, 1988). While it is impossible to know for certain how a particular person will act, prediction is used to reduce uncertainty about a person's future behavior (Clear, 1988; Gottfredson, 1987). Of course, errors are made. The two types of errors associated with prediction are false positives and false negatives. False positives occur when a person is incorrectly predicted to be a risk (i.e., to re-offend, fail on parole/probation, or whatever the criterion may be). False negatives occur when a person is predicted not to be a risk when he or she is a risk. Generally, the ramifications of these errors are that when there is a false positive, the offender is subjected to higher levels of supervision than needed and is treated more harshly by the criminal justice system than is warranted for that person (Clear, 1988). Conversely, false negatives mean that the community is at jeopardy. Further, this type of error can be politically problematic for criminal justice agencies if the public becomes aware of it. Thus, there is a tendency to reduce false negatives rather than false positives (ibid).

There are several factors which influence the accuracy of any prediction instrument. First, the base rate, or how frequently an event (such as recidivism) occurs in the population, influences the accuracy of prediction (Clear, 1984; Gottfredson, 1987). In addition to the base rate, there are several other issues that impact the precision of prediction. These include the reliability and validity of the items used, both criterion and predictor (Gottfredson, 1987). Second, selection ratios impact prediction (Gottfredson, 1987). Third, the composition of the sample is important (Gottfredson, 1987). Fourth, the type of statistical technique should be carefully considered. Finally, the validation of any instrument is essential (Clear, 1988; Gottfredson, 1987). Each of these issues is examined in more detail next.

Problems with criterion variables. The measure of "risk" is the criterion variable- that variable that is being predicted. Exactly what "risk" means, however, is unclear. Risk could refer to risk of violation of probation/parole, it could mean subsequent criminality in general or it could mean subsequent violent offenses (Clear, 1988). Thus, one must first determine what type of risk the instrument is predicting when validating a risk prediction instrument.

Further, once it has been determined what the criterion should be, actually measuring it becomes a problem. For example, many researchers (Clear, 1988; Copas and Marshall, 1998; Jones, 1993; NIC, 1980) have explained that when the risk that one is trying to predict is whether an offender will re-offend, the criterion measure is inadequate. Frequently, the criterion is arrest, conviction or incarceration. However, these measure only the crimes for which offenders have been caught, not all re-offenses. Therefore, what is being predicted-future criminal behavior-is not what is measured. Although self-reported measures may be used as well, it is still likely that not all of the re-offenses are recorded. Similarly, using performance on probation or parole as a criterion may present the same difficulties. That is, only those violations that have been detected are measured.

Additionally, Copas and Marshall (1998) indicate that when re-offenses are used as the criterion, there must be a decision made regarding which re-offenses should be considered failures. For example, should traffic offenses be considered? Should reconvictions based on an offense committed prior to the current offense be included? Certainly these will affect the accuracy of prediction.

Problems with predictor variables. Although the NIC argued that there is a small set of variables that tend to consistently predict risk well (prior criminal history, stability, substance use and employment) there are conflicting results concerning which items predict risk. Some variables have been shown to consistently predict risk (or the measure of risk). These include age, gender, education, and whether a weapon was used during the current offense (Bradshaw, 1987; Hoffman and Beck, 1974; Wright et al., 1984). Some variables have been found to have an inconsistent relationship with outcome measures. For example, Wright et al. found that address changes, number of prior convictions, prior probation periods, prior revocations, drug and alcohol use, and percent of time employed did not predict success among probationers. However, others report that these variables do predict outcomes (see Bradshaw, 1987; Hoffman and Beck, 1974). Additionally, Clear (1988) reports that a common finding in the prediction literature is that the seriousness of the current offense does not predict the probability of a subsequent offense, and indeed, is often found to have a negative relationship (the more serious the current offense, the less likely there is to be a subsequent offense and vice versa). However, Gottfredson (1987) argues that prior record "invariably proves of predictive power" (p. 44); that is, it is a good predictor of risk. Further, Neithercutt (1972) discovered that the type of offense is a good predictor of parole outcome: those who have an offense against person are more likely to be remanded to prison for technical violations than those who have an offense against property. Conversely, Galvin and Polk (1981) found little relationship between parole performance and commitment offense.

There are several potential explanations for these different findings. First, Gottfredson points out that predictor variables are not measured in the same way across all studies. Second, it is possible to include variables that really do not predict risk but actually reflect biases in the criminal justice system (Jones, 1993). For example, ethnicity is a variable that has some controversy surrounding its use as a predictor (Petersilia and Turner, 1987). While it sometimes is found to be a good predictor of risk, it is possible that the reason it predicts risk is not because offenders of a certain race are more likely to re-offend but that they are discriminated against and thus are more likely to be arrested, prosecuted or sentenced. Likewise, while Neithercutt (1972) found that offense type was a good predictor of revocation, he argues that the reason it is a good predictor is that criminal justice agents are more apt to revoke parole for those who have an offense against person than those who have an offense against property, not because those with offenses against person are more likely to violate the conditions of parole. Finally, the conflicting results may also reflect the use of different criterion measures.

Selection ratios. A selection ratio is "the proportion of individuals or events studied and identified by the prediction method as belonging to the outcome classification of interest" (Gottfredson, 1987: 26). Selection ratios determine the point on the risk scale where a person changes from a likely success to a likely failure (Jones, 1993). Thus, if the cutoff scores are changed, prediction may become less accurate (Gottfredson, 1987).

Sample selection. The selection of the appropriate sample is essential to ensure that the sample represents the population. It also helps to ensure that the results remain consistent between the construct sample and the validation sample.

Statistical techniques. Researchers who have conducted comparative studies of the results of several analytical techniques have found that simpler techniques may work as well as more complex techniques (see Jones, 1993). Despite this finding, one should choose whatever technique is appropriate for the data (Gardner, Mulvey, and Shaw, 1995; Gottfredson, 1987; Jones, 1993). Further, log-linear models tend to be more appropriate for criminal justice data (Jones, 1993). Whatever method is chosen, it must be appropriate for the data to yield meaningful and reliable results.

Validation of risk prediction instruments. Finally, there is some controversy concerning the transferability of risk instruments. When advocating for the adoption of risk prediction instruments, the NIC argued that the instruments are transferrable with "minor modifications" (NIC, 1980). They did qualify this by arguing that any risk instrument adopted should be altered to reflect the policy of the agency (NIC, 1980) and should be validated for whatever state adopts it (Clear et al., 1984). Since then, researchers have argued that risk instruments are not transferrable (Clear et al., 1984; Gottfredson, 1987; Wagner and Krausman, 1991). Specifically, those who oppose the adoption of existing instruments argue that the population of probationers and parolees change over time and different jurisdictions have different criminal codes and sentencing procedures. Therefore what works in one place and time may not efficiently predict risk in another place and time. Therefore, any instrument, whether adopted from another location or specifically designed for a particular area, must be validated for use for that particular place at that particular time.

The human factor. We have presented several factors that affect the accuracy of risk prediction. However, there is one other factor that affects the accuracy of risk prediction instruments. This is the "human factor." A risk prediction instrument may be completely valid and accurately predict risk, but if it is not implemented the way it was meant to be implemented, then it becomes invalid. Therefore, it is very important that the risk instrument be filled out thoroughly and accurately.

Chapter Three - Data analysis

This chapter addresses the methods used to begin the validation of the RNA. Additionally, we present the preliminary findings. The implications of these findings are presented in the last chapter.

Methods

Data. The data used for this analysis were obtained from two sources. First, an automated database containing the risk and needs scores from the initial assessment and final reassessment was provided by the New Mexico Probation and Parole Division. Second, staff from the Center for Applied Research and Analysis, Institute for Social Research at the University of New Mexico collected additional data from probation and parole files closed between 1991 and 1996. A stratified random sample of 2136 cases proportionate to the number of cases handled by each of the four regional offices was obtained. A sample of 2051 cases remained after listwise deletion for missing data. This sample was then divided into two equal subsamples, one for construction and one for validation. The subsamples were generated using the random sample procedure in SPSS. The results presented here refer *only* to the construction sample.

Dependent variable. The variable that is being predicted for this study is the number of technical violations while on probation or parole. Technical violations are defined as any infraction of the basic conditions of probation or parole or non-compliance with any special conditions. Most probationer or parolees are given special conditions in addition to the basic conditions. There are 27 types of violations; these are listed in Appendix B. The number of technical violations ranges from 0 to 61 with a mean of 2.60 (sd = 3.85). Approximately 62% of those in the sample had at least one technical violation.

Independent variables. There are a total of 23 items on the RNA. All of them are included as predictor variables. Risk items include the following:

- 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.

The needs items are as follows:

- 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

All of the items on the RNA are weighted; as the weight increases, the presumed risk level increases. If the RNA instrument is valid for this outcome, then each of the items in the RNA should have a positive relationship with the number of technical violations. These variables, their weights and descriptive statistics are presented in Appendix C. Note that two items appear on both the risk assessment and needs assessment: alcohol usage and other drug usage. These are weighted differently, but beyond that there is no difference. The lesson plan for PPO basic training in New Mexico states that the items on each the risk and needs portion should be scored consistently.

Several other variables were chosen as predictors. First, prior offense type was chosen despite the conflicting findings regarding its potential as a predictor. This has been included because specific types of prior offenses may indicate that an offender may be more likely to commit a rules violation. For example, if an offender has prior convictions for a drug offense, it is possible that the client may be more likely to violate the no drugs requirement. Since property offenses are already included in the RNA, they are not repeated. Prior offenses consist of two binary variables: whether the client has any prior convictions for violent offenses and whether the client has any prior convictions for drug offenses. Violent offenses include murder, rape, aggravated and other assaults, child abuse, child abuse resulting in death and spouse abuse. Drug offenses include possession, trafficking and possession of drug paraphernalia. Second, whether there was weapon involved in the current offense is also included. It is hypothesized that a weapon indicates a more serious offender, and thus a client who used a weapon would be more likely to perform poorly on probation/parole. Third, the age of the client at intake (in years) is included as well. This variable has been found to be a good predictor of risk. Younger offenders tend to present a greater risk than older offenders. Fourth, marital status is included. It is hypothesized that clients who are married are more likely to perform well on probation/parole because they have ties to the community. Fifth, whether the client lives with his/her friends is included. It is hypothesized that if the client lives with his/her friends, that he/she is more likely to violate conditions of probation/parole than if the client lives with family members or by themselves. Finally, whether the client is on probation or parole is included. Prior research has found that clients who are on parole perform better than clients on probation despite having the same risk score (Clear, 1988). Other variables that may predict outcome, such as gang affiliation, were not included because the information was not consistently recorded in the files.

Control variables. In addition to the independent variables selected, several variables were included as controls. These include length of time on probation or parole (coded in months), gender, and ethnicity (White, Hispanic and African American). Although both gender and ethnicity have been included as predictors in other studies, there is controversy regarding whether such static indicators should be used. However, we have chosen to include them, but only as control variables as suggested by Jones (1993). The coding and descriptive statistics of all the variables used in this study are presented in Appendix C.

Analytic techniques

There are several ways to indicate whether a risk prediction instrument is performing the way it ought to perform. First, we examine the distribution of cases in each level of supervision based on the computed scores. The proportion of cases in lower and higher levels of supervision should be about equal (Clear, 1988; Wagner and Krausman, 1991). When they are not equal, it indicates that the risk instrument may be overclassifying risk (false positives) or underclassifying risk (false negatives).

Second, we examine the outcomes associated with each level of supervision. Outcomes should be worse (a higher mean number of technical violations) for clients classified into more intense levels of supervision and better outcomes for clients in lower levels of supervision.

Third, we compare the mean number of technical violations for each weighted level of the RNA items. This comparison is more specific than the previous comparison and is intended to assess how each item predicts risk. The mean number of technical violations should increase with each increase in the scoring of each item.

Finally, Poisson regression is used to ascertain the effectiveness of the risk portion of the instrument, the needs portion of the instrument and the instrument overall. Poisson regression is used when the dependent variable is a "count" variable, or a measure of the number of times an event occurs.

Results

Proportion of cases in each level of supervision. The RNA scores tend to classify clients into higher levels of supervision: 58% of clients are classified to maximum supervision while less than 8% are classified to minimum supervision. The proportion of cases in each level of supervision is presented in Table 1. The level of supervision clients were actually assigned to is included as well since the classification based on the RNA score can be overridden. There appears to be a tendency to move clients from higher levels of supervision to medium supervision.

Table 3.1 Proportion of cases in computed and actual supervision levels

Supervision level	Computed supervision level	Actual supervision level
Intensive	N/A	6.4%
Maximum	57.6%	45.5%
Medium	34.9%	41.6%
Minimum	7.5%	6.5%

$p<.001$

Over one-third (31.7%) of the clients were assigned to a supervision level that differs from the one that was computed by the RNA. However, recall that the computed level of supervision does not allow for an assignment into intensive supervision (ISP). Combining the maximum and ISP levels of assigned supervision, the proportion of clients whose classification level differs between the computed and assigned levels is 26.2%. Table 2 illustrates the difference between the computed and assigned levels of supervision, combining the ISP and maximum supervision categories. The assigned level of supervision tends to change to the next level up or down rather than two levels up or down. However, in 3.3% of the cases, the clients moved to a level of supervision that was much higher or lower than the RNA score warranted. Further, as can be seen from the table, more clients were moved to lower levels of supervision than to higher levels of supervision.

Table 3.2 Change in level of supervision

Level change	N	%
Up two levels	13	1.3
Up one level	101	9.9
No change	756	73.8
Down one level	135	13.2
Down two levels	20	2.0

Mean number of technical violations by classification level. The average number of technical violations for each level of supervision, both computed and assigned, is presented next. The mean number of technical violations does decrease for each of the computed classification levels, as would be expected (see Table 3). Overall, there was a statistically significant difference found in the mean number of technical violations for clients in the three levels of supervision. However, the difference between the average number of technical violations is not significant between those classified as medium and minimum security; the difference is significant ($p<.001$) when comparing clients classified to maximum supervision as compared to the other two categories.

We computed the mean number of technical violations for the level of supervision the client was *actually* classified into for comparison. The results are similar to that of the computed levels. However, the mean number of technical violations is actually lower for clients in intensive supervision as compared to those in maximum supervision. Further, the mean number of technical violations for clients in medium and minimum supervision is virtually identical.

Table 3.3 Average number of technical violations by computed and actual level of supervision

Supervision level	Computed	Actual
	Mean TV* (sd)	Mean TV* (sd)
Intensive	N/A	3.35 (3.15)
Maximum	3.27(4.48)	3.44 (4.86)
Medium	1.89 (3.00)	1.82 (2.85)
Minimum	1.49 (3.45)	1.81 (3.34)

$p<.001$, but n/s between intensive/maximum and medium/minimum;
n/s difference between medium and minimum

Risk scores and mean number of technical violations. The mean number of technical violations increased as the scores increased for seven of the eleven items on the risk assessment. There was variation in the mean number of technical violations for the remaining four items: alcohol usage problems, attitude, conviction or juvenile adjudications for various property offenses, and conviction for an assault within the last five years. As can be seen in Table 4, the mean number of technical violations is highest for those who had occasionally abused alcohol as compared to those in the other two categories. Although the differences in the mean number of technical violations was statistically significant overall, a subsequent examination of each pair of means revealed that there is no statistically significant difference in the number of technical violations between those in the two extreme groups. A similar pattern was found with the item related to the client's attitude: the largest number of technical violations was associated with the middle category. Subsequent tests were performed on each possible pair; the results indicate that the difference in the mean number of technical violations is statistically significant between those who are motivated to change and those who are dependent or unwilling to accept responsibility; however, there was no statistically significant difference found between the other two pairs. The mean number of technical violations for prior convictions for property offenses was highest for those who had been convicted of offenses in both categories (worthless checks or forgery and burglary, theft, auto theft or robbery), followed by those who had previously been convicted only for burglary, theft, auto theft or robbery. As might be expected, when subsequent tests were completed, no statistically significant differences were found between those who did not have any convictions for a property offense and those who had prior convictions for worthless checks or forgery. Likewise, no statistically significant difference was found between those who had a burglary, theft, auto theft, or robbery offense and those who had both types of offenses. Finally, clients who had a recent conviction for assault had a lower mean number of technical violations as compared to those who did not have a recent conviction for an assault; however, this difference was not statistically significant. Indeed, this was the only item on the risk assessment portion of the RNA that was not statistically significant.

Table 3.4 Average number of technical violations by risk items

Risk items	**Mean (sd) number of technical violations**	**N**
Address changes***		
0 None	2.31 (3.17)	540
2 One	2.55 (3.84)	295
3 Two or more	3.83 (5.80)	190
Time employed*		
0 60% or more and not applicable	2.34 (4.27)	500
1 40% to 59%	2.86 (3.60)	193
2 Under 40%	3.02 (3.79)	332
Alcohol usage problems**		
0 No interference with functioning	2.25 (3.24)	342
2 Occasional abuse: some disruption of functioning	3.14 (5.08)	343
4 Frequent abuse: serious disruption, needs treatment	2.58 (3.39)	340
Other drug usage problem***		
0 No interference with functioning	1.98 (3.02)	505
1 Occasional abuse: some disruption of functioning	2.87 (4.91)	290
2 Frequent abuse: serious disruption: needs treatment	3.89 (4.34)	230
Attitude**		
0 Motivated to change, receptive to assistance	2.38 (3.98)	242
3 Dependent or unwilling to accept responsibility	3.45 (4.32)	215
5 Rationalizing behavior, negative, not motivated	3.18 (2.79)	68
Age at first adjudication***		
0 24 or older	1.81 (3.07)	392
2 20 to 23	2.98 (5.33)	221
4 19 or younger	3.28 (3.84)	412
Number of prior periods of probation/parole***		
0 None	2.08 (3.30)	578

Risk items	Mean (sd) number of technical violations	N
4 One or more	3.40 (4.67)	447
Number of prior probation/parole revocations***		
0 None	2.28 (3.39)	816
4 One or more	4.11 (5.59)	209
Number of prior felony convictions***		
0 None	2.20 (3.99)	695
2 One	3.19 (3.22)	183
4 Two or more	4.16 (4.55)	147
Conviction or juvenile adjudications for:***		
0 None	2.16 (4.00)	621
2 Burglary, theft, auto theft or robbery	3.55 (4.01)	322
3 Worthless checks or forgery	2.44 (3.46)	55
5 Both categories	3.93 (3.20)	27
Conviction or juvenile adjudication for assault in last 5 years		
0 No	2.67 (4.33)	765
15 Yes	2.61 (2.87)	260

*$p<.05$
**$p<.01$
***$p<.001$

Needs scores and technical violations. Unlike the risk score items, only half of the needs score items were statistically significant (see Table 5). Among those items that were found to have statistically significant differences, the mean number of technical violations did not follow the expected pattern for three of the needs items: academic/vocational skills, employment and alcohol usage. All three follow the same pattern: the last category of each of the items has a lower mean number of technical violations than the category before it. Thus, for academic/vocational skills, the mean number of technical violations for minimal skill level is lower than the mean number of technical violations for clients who are unemployed but have job skills.

Table 3.5 Average number of technical violations by needs items

Needs items	Mean number of technical violations	N
Academic/vocational skills***		
-1 High school or above	1.86 (3.23)	254

Needs items	**Mean number of technical violations**	**N**
0 Adequate skills: able to handle everyday requirements	2.77 (3.59)	510
2 Low skill level causing minor adjustment problems	3.53 (5.55)	216
4 Minimal skill level causing serious adjustment problems	2.76 (2.61)	45
Employment**		
-1 Satisfactory employment for one year or more	1.69 (2.89)	62
0 Secure employment: no difficulties reported	2.27 (4.31)	428
3 Unsatisfactory employment/unemployed but has adequate job skills	3.10 (3.93)	479
6 Unemployed and virtually unemployable	2.89 (2.76)	56
Financial Management		
-1 Long standing pattern of self-sufficiency	1.58 (4.58)	12
0 No current difficulties	2.40 (3.21)	287
3 Situational or minor difficulties	2.79 (4.31)	655
5 Severe difficulties	2.69 (3.94)	71
Marital/family relationships		
-1 Relationships and support exceptionally strong	3.50 (5.53)	8
0 Relatively stable relationships	2.48 (3.44)	544
3 Some disorganization of stress but potential for improvement	2.77 (4.68)	387
5 Major disorganization or stress	3.20 (3.93)	86
Companions***		
-1 Good support and influence	0	4
0 No adverse relationships	1.95 (3.25)	460
2 Associations with occasional negative results	3.05 (4.47)	448
4 Associations almost completely negative	4.05 (4.36)	113
Emotional stability		
-2 Exceptionally well adjusted; accepts responsibility for actions	.75 (.96)	4
0 No symptoms of emotional instability	2.60 (4.09)	850
4 Symptoms limit but do not prohibit adequate functioning	2.94 (3.39)	148
7 Symptoms prohibit adequate functioning	3.39 (4.85)	23
Alcohol usage***		
0 No interference with functioning	2.24 (3.20)	364
3 Occasional abuse; some disruption of functioning	3.29 (5.14)	350
6 Frequent abuse; serious disruption, needs treatment	2.42 (3.25)	311
Other drug usage***		
0 No interference with functioning	2.03 (3.03)	518
3 Occasional abuse, some disruption of functioning	2.91 (4.81)	337
5 Frequent abuse, serious disruption, needs treatment	4.08 (4.46)	170
Mental health	2.66 (4.05)	975
0 Able to function independently	2.70 (3.10)	46
3 Some need for assistance, potential for adequate adjustment	2.25 (2.06)	4
6 Deficiencies severely limit independent functioning	2.67 (4.06)	943
Physical Health	2.78 (3.92)	55
0 Sound physical health, seldom ill	1.81 (2.08)	27
1 Handicap or illness interferes with functioning on regular basis		
2 Serious handicap or chronic illness	2.66 (3.98)	979
Sexual behavior	2.79 (5.20)	21
0 No apparent dysfunction	2.32 (4.16)	25
3 Real or perceived situational or minor problems		
5 Real or perceived chronic or severe problems	.33 (.57)	3

Needs items	Mean number of technical violations	N
PPO's impression of client's needs*** -1 Minimum 0 Low 3 Medium 5 Maximum	1.45 (2.36) 2.35 (4.23) 3.21 (3.91)	78 493 451

*$p<.05$
**$p<.01$
***$p<.001$

Multivariate analysis. Four models were generated using Poisson regression to test three hypotheses. The first model includes only those items from the risk portion of the RNA. The second model contains only those items from the needs portion of the RNA. The third model contains all of the items from the RNA. The fourth model includes all of the items on the RNA in addition to other independent variables and control variables that were previously discussed.

The first hypothesis examined tests whether the needs portion of the RNA significantly improves the fit of the model. In other words, does the needs assessment predict technical violations? By comparing model one (risk items only) with model three (both risk and needs items), this hypothesis can be tested. The addition of the needs items does significantly improve fit. Next, we compare model two to model three. This comparison tests whether including the risk items in the RNA significantly improves the fit of the model. Again, this is statistically significant revealing that the inclusion of the risk items does improve the fit of the model. We cannot compare the likelihood ratios to determine whether model one (risk only) or model two (needs only) is better since the two models are not nested (all of the variables in one model must be in the other model as well if the models are nested). However, we can examine the R^2 which can be used to assess predictive efficacy of models. The R^2 is larger for model one as compared to model two, and similar to that of model three. This suggests that including the needs assessment items does not improve the predictive efficacy of the instrument much. However, since the likelihood ratio does indicate statistical significance, the needs assessment items have some statistical value. In order to determine which particular items predict risk, we examine each of the regression coefficients next. Since model three which includes all the RNA items was a significant improvement over each of the other two models, we focus on those coefficients.

<u>*Coefficients related to risk items*</u>. Six items from the risk assessment-*number of prior periods of probation/parole, prior convictions for property offenses* and *convictions for assault* within the last five years- were not statistically significant. Three other items (*time employed*, *alcohol usage* and *attitude*) are marginally significant ($\underline{p}<.05$). The percentage of *time employed* has a negative relationship with technical violations indicating that as the percentage of time employed increases, the number of technical violations increases. This is contrary to what would be expected if the variable is a good predictor of risk. However, this is consistent with the bivariate analysis. The sign for *alcohol usage* is in the direction expected for model three. Interestingly, however, there was an inverse relationship found for this coefficient in model one and it was not statistically significant. The coefficient for *attitude* was in the direction expected. The number of *prior periods of probation/parole* was not statistically significant, but *revocations* was significant ($\underline{p}<.01$) indicating that past performance on probation/parole predicts subsequent performance on probation/parole. In addition, four other items on the risk assessment are statistically significant and positively associated with the number of technical violations. These include the number of *address changes*, *drug use problems, age at first adjudication,* and the *number of prior felony convictions*.

<u>*Coefficients related to needs items*</u>. Many of the items on the needs assessment portion of the RNA had an inverse relationship with the number of technical violations, but most of these were not statistically significant. Two of the twelve needs assessment items were marginally significant: *mental health* and the *PPOs impression of the client's needs*. The coefficient corresponding to the *mental health* of the client has an inverse relationship with the number of technical violations, contrary to expectations. The statistical significance of the *PPOs impression of client's needs* decreases between model two and model three, suggesting that once the risk items are taken into account, this variable is not as important in predicting risk of technical violations. Four items on the needs assessment were statistically significant ($\underline{p}<.01$); two of these were positively related to the number of technical violations. A decrease in

academic/vocational skills was related to an increase in technical violations. Recall that the percentage of time currently employed from the risk assessment portion was only marginally significant. The adequacy of the client's employment from the needs assessment portion is not statistically significant. This may imply that the skills a person has is more important when predicting risk than their employment situation.

An increase in the number of negative *companions* also increases the number of technical violations. Two other items from the needs assessment were statistically significant ($p<.01$), but were inversely related to technical violations: marital/family relationships and alcohol use. Increases in the instability of *marital/family relationships* correspond to a decrease in the number of technical violations. *Alcohol usage problems* also has an inverse relationship with technical violations. Both drug use and alcohol use are redundant items on the two parts of the RNA. The alcohol usage item on the risk portion was found to have a marginally significant positive relationship with the number of technical violations, but the sign changed from model two to model three. The same item on the needs portion has a statistically significant inverse relationship with risk. The only difference between the two items is the weighting. Further, drug use was statistically significant when used in the risk assessment, but is not statistically significant in the needs assessment. It is possible that statistical significance was found for only one of the items because these items are highly correlated, causing the standard errors to inflate.[1]

Table 3.6 Poisson regression results for model one, two and three

Independent Variable	Model one (Risk scores only)	Model two (Need scores only)	Model three (Full RNA)	Factor change in expected count[2]
Constant	.358 (.045)**	.386 (.065)**	.289 (.068)**	
Address changes	.105 (.016)**		.112 (.016)**	1.12

[1]The correlation between the two alcohol usage items and the two drug usage items was computed. The correlations were not equal to one; rather they were .92 and .90, suggesting that those who complete the forms are not always consistent.

[2]The factor change is based on the coefficients produced in Model 3; only those coefficients that are statistically significant are included.

Time employed	-.025 (.023)		-.064 (.025)*	0.94
Alcohol usage problems	-.009 (.013)		.065 (.029)*	1.07
Other drug usage problems	.221 (.023)**		.145 (.051)**	1.16
Attitude	.039 (.012)**		.028 (.012)*	1.03
Age at first adjudication	.074 (.013)**		.060 (.013)**	1.06
Number of prior periods of probation/parole	.007 (.014)		.004 (.014)	
Number of prior probation/parole revocations	.049 (.014)**		.039 (.014)**	1.04
Number of prior felony convictions	.038 (.017)*		.045 (.017)**	1.05
Prior convictions for property offenses	.039 (.017)*		.028 (.017)	
Conviction for assault in last 5 years	-.003 (.003)		-.005 (.003)	
Academic/vocational skills		.072 (.016)**	.085 (.016)**	1.09
Employment		.014 (.012)	.003 (.013)	
Financial Management		-.021 (.014)	-.019 (.014)	
Marital/family relationships		-.027 (.012)*	-.040 (.013)**	0.96
Companions		.114 (.016)**	.089 (.016)**	1.09
Emotional stability		.024 (.012)*	.024 (.012)	
Alcohol usage		-.020 (.008)*	-.059 (.019)**	0.94
Other drug usage		.086 (.011)**	.013 (.021)	
Mental health		-.066 (.029)*	-.070 (.029)*	0.93
Physical Health		-.103 (.055)	-.054 (.056)	
Sexual behavior		-.015 (.023)	-.004 (.024)	
PPOs impression of needs		.089 (.017)**	.039 (.017)*	1.04
Number of observations	1025	1025	1025	
Likelihood ratio	4004.30	4094.95	3913.65	
Pseudo-R^2	.022	.004	.025	

<u>*Inclusion of variables not currently used in the RNA*</u>. The last hypothesis questions whether the items included on the RNA sufficiently predict risk or whether other items should be included. A final model compares model three from the previous set of nested models to a fourth model which includes all of the items currently used in the RNA along with nine other items. The results are shown in Table 7. The likelihood ratios are compared to assess whether the fit of the model improves with the addition of these other variables. This is statistically significant indicating that including these variables does improve the prediction of the number of technical violations. The new items include both independent and control variables. These are discussed next.

<u>*Additional independent variables*</u>. One *prior offense type* was marginally significant: whether

the client had a previous conviction for a drug offense. Whether the client had a previous conviction for a violent offense was not statistically significant, but if a *weapon* was used during the commission of the current offense was significant ($p<.01$). The number of technical violations increases by approximately 24% when a weapon was used, holding all other variables constant. The client's *age* at intake has a negative coefficient, as anticipated, and is statistically significant ($p<.01$). The coefficient associated with *marital status* is also negative which was expected, and is marginally significant ($p<.05$). Whether the client was *living with friends* was statistically significant and had a positive coefficient. The *probation versus parole status* was not statistically significant relationship.

Control variables. Two of the control variables were statistically significant: the length of time the client had been on probation/parole and one of the variables measuring ethnicity. The longer clients had been on probation/parole, the more technical violations they had, as would be expected. Only one category of ethnicity was statistically significant: whether the client was African American. The remaining two categories of ethnicity were not statistically significant, nor was the gender of the client.

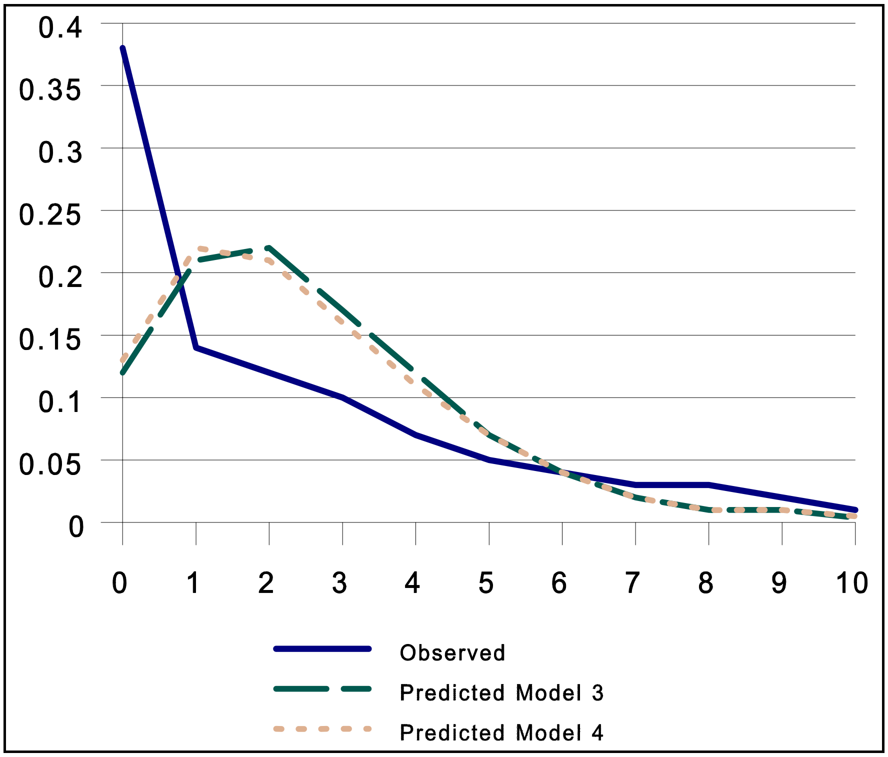
RNA items. Except for the coefficient for the physical health of the client, which was not statistically significant, the sign of the coefficients for all of the RNA items remained the same from model three to model four. Two of the coefficients became statistically significant in model four, including whether there was a *conviction for an assault* within the last five years ($p \leq .01$) and the *emotional stability* of the client ($p \leq .05$). Interestingly, the PPOs impression of the client's needs was statistically significant ($p \leq .01$) in model two, is less significant in model three, and is not significant in model four.

Table 3.7 Poisson regression results for models three and four

Independent Variable	Model 3 (Full RNA)	Model 4 (RNA with other variables)
Constant	.289 (.068)**	.459 (.154)**
Address changes	.112 (.016)**	.085 (.016)**
Time employed	-.064 (.025)*	-.077 (.026)**
Alcohol usage problems	.065 (.029)*	.078 (.029) **
Other drug usage problems	.145 (.051)**	.126 (.051)*
Attitude	.028 (.012)*	.040 (.013)**
Age at first adjudication	.060 (.013)**	.031 (.015)*
Number of prior periods of probation/parole	.004 (.014)	.008 (.015)
Number of prior probation/parole revocations	.039 (.014)**	.039 (.014)**
Number of prior felony convictions	.045 (.017)**	.067 (.018)**
Prior convictions for property offenses	.028 (.017)	.023 (.017)
Conviction for assault in last 5 years	-.005 (.003)	-.010 (.004)**
Academic/vocational skills	.085 (.016)**	.071 (.016)**
Employment	.003 (.013)	.009 (.013)
Financial Management	-.019 (.014)	-.022 (.014)
Marital/family relationships	-.040 (.013)**	-.039 (.013)**
Companions	.089 (.016)**	.084 (.017)**
Emotional stability	.024 (.012)	.028 (.013)*
Alcohol usage	-.059 (.019)**	-.052 (.020)**
Other drug usage	.013 (.021)	.017 (.022)
Mental health	-.070 (.029)*	-.078 (.030)**
Physical Health	-.054 (.056)	.009 (.057)
Sexual behavior	-.004 (.024)	-.003 (.024)
PPOs impression of needs	.039 (.017)*	.031 (.017)
Prior offense types		
Prior convictions for a violent offense		.066 (.053)
Prior convictions for a drug offense		.104 (.051)*
Weapon used during commission of current offense		.218 (.057)**
Age at intake		-.012 (.003)**
Married		-.115 (.049)*
Living with friends		.446 (.075)**
Probationer not parolee		.025 (.057)
Length of time in probation/parole (in months)		.007 (.002)**
Male client		.012 (.059)
Ethnicity		
White client		.021 (.089)
Hispanic client		-.003 (.084)
African American client		.407 (.102)**
Number of observations	1025	1025
Likelihood ratio	3913.65	3771.42
Pseudo-R^2	.025	.075

__Predicted probability of technical violations__. Finally, to illustrate how well the models predict the number of technical violations, we computed the mean predicted probability of zero to ten technical violations for both model three and model four, based on the Poisson regression coefficients. This is compared to the observed proportion of technical violations, as seen in Figure 1. Neither model three nor model four closely follow the observed pattern of technical violations. Both tend to underpredict the number of zeros, overpredict the number of ones through fives and underpredict the number of sevens through nines. Long (1997) explains that

when there are large differences between the mean probabilities and observed proportions, the model is inappropriate. This indicates that neither model three nor model four are appropriate for this data.



Chapter Four - Discussion

Discussion

There are several tentative conclusions that can be drawn based on the findings presented in this report. First, we pointed out that there is not a cutoff score for classifying offenders into ISP. Thus, clients are assigned to ISP, but the RNA score does not determine whether they are accepted into that program. Further, the supervision level of clients in other special case management programs is not determined by the RNA scores. This suggests that the RNA should be altered to address these programs.

Second, we found that the greatest proportion of clients were classified into the maximum risk category when the RNA instrument is used. This could indicate that the instrument tends to overpredict risk of technical violations. Additionally, we found that when clients were assigned to a level that differed from the one that was computed by the RNA, more clients were moved into lower levels of supervision rather than higher. There are at least two possible reasons for the movement of clients into lower levels of supervision. First, one plausible explanation is that the RNA tends to classify clients into higher levels of supervision than is needed as suggested by the proportion of clients in each computed supervision level. An alternative argument is that the PPOs move clients into lower levels of supervision so that they are not required to supervise the clients as closely. It could be, perhaps, that it is a combination of these reasons that account for the movement of clients into lower levels of supervision.

Next, when examining the mean number of technical violations for each level of supervision computed by the RNA instrument we discovered that those clients who were calculated as maximum risk had the greatest number of technical violations as compared to those who were determined to be medium or minimum risk. However, there was no statistically significant difference found in the mean number of technical violations between those clients who were classified as medium risk and those who were minimum. This finding suggests that while the RNA instrument may predict risk well for those who are maximum risk, it is less able to differentiate between those who are determined to be a lesser risk.

Fourth, we examined the mean number of technical violations for each risk item individually. We determined that seven of the items on the risk portion of the RNA followed the pattern that would be expected if these items and each of their response categories predict risk of technical violations. Specifically, we discovered that the number of technical violations increased with each category for the following variables: address changes, time employed, drug usage problems, age at first adjudication, number of prior periods of probation/parole, number of prior probation/parole revocations, and number of prior felony convictions. The remaining four items (alcohol usage problems, attitude, conviction for a property offense and conviction for a recent assault) did not follow the expected pattern. Only one risk item- recent conviction for assault- was not statistically significant.

Each of the items on the needs portion was examined as well. Only half of the items were statistically significant. Among those, only three items- companions, drug usage, and PPOs impression of client's needs, followed the expected pattern. The remaining items did not. One observation that can be made is that many of the extreme categories have few cases. For example, less than 5% of cases are categorized as having a minimal educational skill level. This may help to explain why the mean number of technical violations does not always follow the pattern it should. It is possible that these extreme categories should be combined into other categories. For example, it may be sensible to combine low and minimum skills for the item that measures academic/vocational skills.

We then presented a series of multivariate Poisson models. The results suggest that while some items on the RNA appear to predict risk, there are items that do not. Generally, the risk assessment portion of the RNA seems to predict risk better than the needs portion, as indicated in the bivariate analysis. To reiterate, the items that appear to predict the number of technical violations, one measure of performance while on probation or parole, are as follows:

- number of address changes

- length of time employed
- alcohol usage problems
- other drug usage problems
- attitude
- age at first adjudication
- number of prior probation/parole revocations
- number of prior felony convictions
- academic/vocational skills
- marital/family relationships
- companions
- mental health

In addition, two items, emotional stability and PPOs impression of needs, may have a relationship with risk. Several items that are not currently used on the RNA were included in the last model and were found to significantly predict risk. These items include the following:

- prior convictions for a drug offense
- weapon used during the commission of the current offense
- age at intake
- whether the client was married
- whether the client was living with friends
- length of time in probation/parole

The inverse relationships found between some of the items on the RNA and the number of technical violations suggests several possibilities. First, it could be that the items may not be appropriate for determining risk level. Conversely, it may imply that the items may need to be altered somewhat in terms of the categories included. For example, in model Four one of the new items, marital status, is somewhat comparable to the needs item, strength of family/marital relationships. One would expect that both items would have a similar relationship to technical violations. That is, a client who is married should have fewer violations (which the coefficient shows) and that a client with strong family/marital relationships should have fewer technical violations (which the coefficient does not support). It is counter-intuitive that these results are different. One possible reason for this is that the categories for the strength of marital/family relationships may be too specific. If the variable were changed to two categories-relatively stable relationships or relatively unstable relationships- it may predict risk better. Finally, it could be that the weighting is inappropriate. These possibilities indicate that further analyses should be conducted. A multivariate analysis using the categories of each of the items as dummy variables (essentially, each category becomes a variable) would help to determine whether the categories should be revised and/or whether the weighting should be changed. Based on these results, a revised risk/needs instrument could be devised. In order to determine whether this newer instrument has better predictive power, it would have to be validated using the other half of the sample which was not used here.

Additionally, it was concluded that none of the models really fit the data well. Long (1997) explains that Poisson regression seldom fits the data because there is overdispersion. This means that the variance is greater than the mean. The primary problem overdispersion causes is that the standard errors of the estimates are biased: they are smaller than they should be (Barron, 1992; Land, McCall and Nagin, 1996). This results in coefficients that are statistically significant, but should not be significant (or at least the coefficients appear more significant than they really are). In other words, the Poisson regression indicates that there are variables that predict the outcome measure with statistical significance but really should not predict the outcome. One way this problem is eliminated is when all of the variables that predict the dependent variable are included. However, this is unlikely. A second way to overcome the overdispersion problem is to use a negative binomial regression which lifts that particular restriction: that is, the variance can exceed the mean. Therefore, a negative binomial regression should be computed to determine whether the poor model fit is in part due to overdispersion. It is very likely that the difference in the results between the Poisson and the negative binomial regression is that fewer RNA items would be statistically significant.

We noted that the correlation between the alcohol and drug items on the risk and needs portions of the RNA was high but not perfect. This suggests that the forms are not being filled in accurately. Additionally, a survey of the probation and parole officers conducted by the ISR indicates that less than half (46%) of probation and parole officers think that the risk assessment form is at least somewhat helpful in performing their job duties. Further, only 36% of probation and parole officers feel that the scoring procedures for the RNA gives an accurate picture of the client's risk. This suggests that the majority of probation and parole officers do not feel that the RNA is useful or accurate. This *may* mean that less care is taken when filling out these forms. If this is true, the problem this creates is twofold. First, as noted previously, the best risk prediction device is only effective if it is used the way it was meant to be used. Therefore, even if the RNA itself is valid, if it is not used appropriately it invalidates the instrument. Second, the validation of the RNA is based on the use of the instrument. If the instrument is filled in appropriately, the results of the validation check are sound; however, if the instrument is systematically filled in inappropriately, the results of the validation check are questionable. We have to assume that while there are some errors, the majority of the information is accurate.

Future analysis

The next series of steps towards validating the RNA instrument are as follows. First, we will conduct a negative binomial regression to determine whether this type of regression would be more appropriate for this data. Next, we will conduct a dummy regression to help determine whether the categories used for the RNA items should be combined and whether different weighting should be used. Third, we will conduct a logistic regression to determine whether the RNA items predict the clients' ultimate termination status (completed or not) from probation or parole. These analyses will help to determine whether the RNA is valid for predicting how well clients do on probation or parole. However, it will not be validated for predicting recidivism until we receive the subsequent arrest information.

Appendix A. Cutoff scores

Table A.1 Original score cutoffs used by Wisconsin

Supervision level	Risk score	Need score
Maximum	15+	30+
Medium	8 to 14	15 to 29
Minimum	0 to 7	-7 to 14

Source: NIC, 1980

Table A.2 New Mexico's cutoff scores

Supervision level	Risk score	Need score
Maximum	15+	20+
Medium	8 to 14	8 to 19
Minimum	0 to 7	-7 to 7

Appendix B. Technical violations.

Table B.1 Description of technical violations

Type of violation
Violation of laws.
Failure to report to probation/parole officer.
Failure to get permission to leave the county, change jobs, change residence, enter into civil contracts or enrolling/withdrawing from school.
Associating with persons having a criminal record or anyone determined by the probation/parole officer (PPO) as a detriment to probation/parole.
Failure to follow orders/instruction of probation/parole or failure to promptly respond to correspondence from PPO.
Failure to permit PPO to visit with client at reasonable times and places.
Failure to obtain employment.
Possessing a weapon.
Using a controlled substance or alcohol to excess; failure to provide urine specimens to PPO.
Failure to report a new arrest within 72 hours of the incident.
Failure to pay probation/parole costs.
Failure to receive permission to act as an informer for a law enforcement agency.
Failure to enter/complete treatment
Failure to pay restitution
Failure to pay crime stoppers fee/community programs fee.
Failure to pay child support
Failure to pay lab fee/intoximeter fee
Failure to complete community service
Failure to maintain no contact with victim(s)
Failure to maintain no contact with children
Failure to meet curfew
Failure to appear for hearing
Failure to abstain from alcohol or stay out of bars
Failure to attend AA/NA meetings
Failure to attend alcohol/drug screening
Failure to attend/complete domestic violence counseling
Other

Appendix C. Coding and descriptive statistics for dependent, independent and control variables

Table C.1 Variables included in analysis

Variables included in analysis	**%**	**Mean (sd)**
Dependent variable		
Technical violations		**2.66 (4.01)**
Independent variables		
Risk items		
Address changes		1.13 (1.24)
0 None	52.7	
2 One	28.8	
3 Two or more	18.5	
Time employed		.84 (.89)
0 60% or more	48.8	
1 40% to 59%	18.8	
2 Under 40%	32.4	
0 Not applicable		2.00 (1.63)
Alcohol usage problems		
0 No interference with functioning	33.4	
2 Occasional abuse: some disruption of functioning	33.5	
4 Frequent abuse: serious disruption, needs treatment	33.2	
Other drug usage problem		.73 (.80)
0 No interference with functioning	49.3	
1 Occasional abuse: some disruption of functioning	28.3	
2 Frequent abuse: serious disruption: needs treatment	22.4	
Attitude		.96 (1.62)
0 Motivated to change, receptive to assistance	72.4	
3 Dependent or unwilling to accept responsibility	21.0	
5 Rationalizing behavior, negative, not motivated	6.6	
Age at first adjudication		2.04 (1.77)
0 24 or older	38.2	
2 20 to 23	21.6	
4 19 or younger	40.2	
Number of prior periods of probation/parole		1.74 (1.98)
0 None	56.4	
4 One or more	43.6	
Number of prior probation/parole revocations		.82 (1.61)
0 None	79.6	
4 One or more	20.4	
Number of prior felony convictions		.93 (1.46)
0 None	67.8	
2 One	17.9	
4 Two or more	14.3	
Conviction or juvenile adjudications for:		.92 (1.25)
0 None	60.6	
2 Burglary, theft, auto theft or robbery	31.4	
3 Worthless checks or forgery	5.4	
4 Both categories	2.6	

Variables included in analysis	**%**	**Mean (sd)**
Conviction or juvenile adjudication for assault in last 5 years		3.80 (6.53)
0 No	74.6	
15 Yes	25.4	
Needs items		
Academic/vocational skills		.35 (1.29)
-1 High school or above	24.8	
0 Adequate skills: able to handle everyday requirements	49.8	
2 Low skill level causing minor adjustment problems	21.1	
4 Minimal skill level causing serious adjustment problems	4.4	
Employment		1.67 (1.86)
-1 Satisfactory employment for one year or more	6.0	
0 Secure employment: no difficulties reported	41.8	
3 Unsatisfactory employment/unemployed but has adequate job skills	46.7	
6 Unemployed and virtually unemployable	5.5	
Financial Management		2.25 (1.56)
-1 Long standing pattern of self-sufficiency	1.2	
0 No current difficulties	28.0	
3 Situational or minor difficulties	63.9	
5 Severe difficulties	6.9	
Marital/family relationships		1.54 (1.77)
-1 Relationships and support exceptionally strong	.8	
0 Relatively stable relationships	53.1	
3 Some disorganization of stress but potential for improvement	37.8	
5 Major disorganization or stress	8.4	
Companions		1.31 (1.34)
-1 Good support and influence	.4	
0 No adverse relationships	44.9	
2 Associations with occasional negative results	43.7	
4 Associations almost completely negative	11.0	
Emotional stability		.73 (1.70)
-2 Exceptionally well adjusted; accepts responsibility for actions	.4	
0 No symptoms of emotional instability	82.9	
4 Symptoms limit but do not prohibit adequate functioning	14.4	
7 Symptoms prohibit adequate functioning	2.2	
Alcohol usage		2.84 (2.43)
0 No interference with functioning	35.5	
3 Occasional abuse; some disruption of functioning	34.1	
6 Frequent abuse; serious disruption, needs treatment	30.3	
		1.82 (1.95)
Other drug usage	50.5	
0 No interference with functioning	32.9	
3 Occasional abuse, some disruption of functioning	16.6	
5 Frequent abuse, serious disruption, needs treatment		.16 (.72)
Mental ability	95.1	
0 Able to function independently	4.5	
3 Some need for assistance, potential for adequate adjustment	.4	
6 Deficiencies severely limit independent functioning		
Health		.11 (.38)
0 Sound physical health, seldom ill	92.0	

Variables included in analysis	%	Mean (sd)
1 Handicap or illness interferes with functioning on regular basis	5.4	
2 Serious handicap or chronic illness	2.6	
Sexual behavior		.18 (.87)
0 No apparent dysfunction	95.5	
3 Real or perceived situational or minor problems	2.0	
5 Real or perceived chronic or severe problems	2.4	
PPO's impression of client's needs		3.64 (1.44)
-1 Minimum	.3	
0 Low	7.6	
3 Medium	48.1	
5 Maximum	44.0	
Other independent variables		
Prior offense types		
Ever convicted of a violent offense?		.17 (.38)
0 No	82.9	
1 Yes	17.1	
Ever convicted of a drug offense?		.16 (.37)
0 No	83.6	
1 Yes	16.4	
Was a weapon used during the commission of the current offense?		.15 (.36)
0 No	84.8	
1 Yes	15.2	
Age at beginning of probation/parole		31.29 (9.87)
Is client married?		.26 (.44)
0 No	73.6	
1 Yes	26.4	
Does client live with friends?		.05 (.22)
0 No	94.9	
1 Yes	5.1	
Is client on probation or parole?		.84 (.37)
0 Parole	16.3	
1 Probation	83.7	
Control variables		
Length of time on probation/parole (in months)		18.55 (12.58)
Gender		.16 (.36)
0 Male	84.3	
1 Female	15.7	
Ethnicity		
White		.31 (.46)
0 No	69.2	
1 Yes	30.8	
Hispanic		.54 (.50)
0 No	45.7	
1 Yes	54.3	
African American		.06 (.25)
0 No	93.6	
1 Yes	6.4	

Bibliography

Barron, David N. 1992. The analysis of count data: overdispersion and autocorrelation. *Sociological Methodology* 22: 179-220.

Bradshaw, Richard Alfred. 1987. *Multivariate actuarial prediction of felonious recidivism of male parolees: Development and cross-validation of a series of risk assessment models using stepwise logistic regression.* University Microfilms International: Ann Arbor Michigan.

Clear, Todd. 1988. Statistical prediction in corrections. *Research in Corrections* 1: 1-40.

Copas, John and Peter Marshall. 1998. The offender group reconviction scale: A statistical reconviction score for use be probation officers. *Applied Statistics* 47: 159-171.

Galvin, James L. and Kenneth Polk. 1981. Parole guidelines: Suggested research questions. *Crime & Delin*

Gardner, William, Edward P. Mulvey and Esther C. Shaw. 1995. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin* 118: 392-404.

Gottfredson, Stephen D. 1987. Prediction: An overview of selected methodological issues. In Gottfredson, D. M. and Tonry, M. (eds.) *Prediction and Classification: Criminal Justice Decision Making.* Chicago: University of Chicago Press.

Hoffman, Peter B. and James L. Beck. 1974. Parole decision making: A salient factor score. *Journal of Cri*

Jones, Peter R. 1993. *Risk Prediction in Criminal Justice*. Paper prepared for the international Association of Residential and Community Alternatives. Philadelphia PA

Land, Kenneth C., Patricia L. McCall and Daniel S. Nagin. 1996. A comparison of Poisson, negative binomial, and semiparametric mixed Poisson regression models. *Sociological Methods & Research* 24: 387-442.

Liedka, Ray. 1998. Class notes from Fall Sociology 582 course. University of New Mexico.

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage Publications.

Neithercutt, M. G. 1972. Parole violation patterns and commitment offense. *Journal of Research in Crime and Delinquency* 9: 87-99.

NIC. 1980. *Classification in Probation and Parole: A Model Systems Approach*. Washington D.C.: National Institute of Corrections.

Petersilia, Joan and Susan Turner. 1987. Guideline-based justice: Prediction and racial minorities. In Gottfredson, D. M. and Tonry, M. (eds.) *Prediction and Classification: Criminal Justice Decision Making.* Chicago: University of Chicago Press.

Wagner, Dennis and Sue Krausman. 1991. *The Iowa Risk Assessment Study*. Unpublished report, National Council on Crime and Delinquency.

Wright, Kevin N., Clear, Todd R., and Paul Dickson. 1984. Universal applicability of probation risk-assessment instruments. *Criminology* 22: 113-134.